

Suresh Chandra Satapathy
Yogesh Chandra Bhatt
Amit Joshi
Durgesh Kumar Mishra *Editors*

Proceedings of the International Congress on Information and Communication Technology

ICICT 2015, Volume 1

Advances in Intelligent Systems and Computing

Volume 438

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba

e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK

e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary

e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA

e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan

e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia

e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico

e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil

e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland

e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong

e-mail: jwang@mae.cuhk.edu.hk

More information about this series at <http://www.springer.com/series/11156>

Suresh Chandra Satapathy
Yogesh Chandra Bhatt · Amit Joshi
Durgesh Kumar Mishra
Editors

Proceedings of the International Congress on Information and Communication Technology

ICICT 2015, Volume 1

Editors

Suresh Chandra Satapathy
Department of Computer Science
and Engineering
Anil Neerukonda Institute of Technology
and Sciences
Visakhapatnam, Andhra Pradesh
India

Yogesh Chandra Bhatt
College of Technology and Engineering
Maharana Pratap University of Agriculture
and Technology
Udaipur, Rajasthan
India

Amit Joshi
Sabar Institute of Technology
Sabarkantha, Gujarat
India

Durgesh Kumar Mishra
Sri Aurobindo Institute of Technology
Indore, Madhya Pradesh
India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-10-0766-8 ISBN 978-981-10-0767-5 (eBook)
DOI 10.1007/978-981-10-0767-5

Library of Congress Control Number: 2016933209

© Springer Science+Business Media Singapore 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Preface

This AISC volume contains the papers presented at the ICICT 2015: International Congress on Information and Communication Technology. The conference was held during October 9 and 10, 2015 at Hotel Golden Tulip, Udaipur, India and organized by CSI Udaipur Chapter, Division IV, SIG-WNS, SIG-e-Agriculture in association with ACM Udaipur Professional Chapter, The Institution of Engineers (India), Udaipur Local Centre and Mining Engineers Association of India, Rajasthan Udaipur Chapter. It has targeted state-of-the-art as well as emerging topics pertaining to ICT and effective strategies for its implementation of engineering and managerial applications. The objective of this international conference is to provide opportunities for the researchers, academicians, industry persons, and students to interact and exchange ideas, experience and expertise in the current trends and the strategies for information and communication technologies. Besides this, participants will also be enlightened about vast avenues, current and emerging technological developments in the field of ICT in this era, and its applications being thoroughly explored and discussed. The conference has attracted a large number of high-quality submissions and stimulated the cutting-edge research discussions among many academic pioneering researchers, scientists, industrial engineers, and students all over the world and also has provided a forum to researcher. The goals of this forum, as follows, are to: Propose new technologies, share their experiences and discuss future solutions for design infrastructure for ICT; provide common platform for academic pioneering researchers, scientists, engineers, and students to share their views and achievements; enrich technocrats and academicians by presenting their innovative and constructive ideas; focus on innovative issues at international level by bringing together the experts from different countries. Research submissions in various advanced technology areas were received and after a rigorous peer-review process with the help of program committee members and external reviewer, 135 (Vol-I: 68, Vol-II: 67) papers were accepted with an acceptance ratio of 0.43. The conference featured many distinguished personalities like Dr. L.V. Murlikrishna Reddy, President, The Institution of Engineers (India); Dr. Aynur Unal, Stanford University, USA; Ms. Mercy Bere, Polytechnique of

Namibia, Namibia; Dr. Anirban Basu (Vice President and President Elect) Zera GmbH, Germany; Dr. Mukesh Kumar, TITS, Bhiwani; Dr. Vipin Tyagi, Jaypee University, Guna; Dr. Durgesh Kumar Mishra, Chairman Division IV CSI; Dr. Basant Tiwari, and many more. Separate Invited talks were organized in industrial and academia tracks on both days. The conference also hosted few tutorials and workshops for the benefit of participants. We are indebted to ACM Udaipur Professional Chapter, The Institution of Engineers (India), Udaipur Local Centre and Mining Engineers Association of India, Rajasthan Udaipur Chapter for their immense support to make this Congress possible in such a grand scale. A total of 15 Sessions were organized as a part of *ICICT 2015* including 12 technical, one plenary and one Inaugural Session and one Valedictory Session. A total of 118 papers were presented in 12 technical sessions with high discussion insights. The total number of accepted submissions was 139 with a focal point on ICT. The Session Chairs for the technical sessions were Dr. Chirag Thaker, GEC, Bhavnagar, India; Dr. Vipin Tyagi, Jaypee University, MP, India; Dr. Priyanka Sharma, Raksha Shakti University, Ahmedabad, India; Dr. S.K. Sharma; Dr. Bharat Singh Deora; Dr. Nitika Vats Doohan, Indore; Dr. Mahipal Singh Deora; Dr. Tarun Shrimali; and Dr. L.C. Bishnoi.

Our sincere thanks to all sponsors, press, and print and electronic media for their excellent coverage of this congress.

October 2015

Suresh Chandra Satapathy
Yogesh Chandra Bhatt
Amit Joshi
Durgesh Kumar Mishra

Organising Committee

Chief Patron

Prof. Bipin V. Mehta, President, CSI

International Advisory Committee

Chandana Unnithan, Victoria University, Melbourne, Australia

Dr. Aynur Unal, Standford University, USA

Dr. Malay Nayak, Director-IT, London

Chih-Heng Ke, MIEEE, NKIT, Taiwan

Dr. Pawan Lingras, Professor, Saint Mary University, Canada

National Advisory Committee

Dr. Anirban Basu, Vice President, CSI

Mr. Sanjay Mahapatra, Hon. Secretary, CSI

Mr. R.K. Vyas, Treasurer, CSI

Mr. H.R. Mohan, Immediate Past President, CSI

Prof. Vipin Tyagi, RVP III, CSI

Prof. S.S. Sarangdevot, VC, JRVU, Udaipur

Dr. R.S. Shekhawat, RSC, Region 3, CSI

Prof. H.R. Vishwakarma, VIT, Vellore, India

Prof. Dr. P. Thrimurthy, Past President, CSI

Dr. G.P. Sharma, CSI Udaipur Chapter

Dr. Nilesh Modi, Chairman, CSI Ahmedabad Chapter

Prof. R.C. Purohit, CSI Udaipur Chapter

Prof. S.K. Sharma Director PIE, PAHER, Udaipur

Dr. T.V. Gopal, Anna University, Chennai

Dr. Deepak Sharma, CSI Udaipur Chapter
Prof. R.C. Purohit, CTAE, Udaipur
Prof. Pravesh Bhadviya, Director, Sabar Education, Gujarat
Dr. Bharat Singh Deora, Department of CS and IT, JRN RV University, Udaipur

Technical Program Committee

Chair

Dr. Suresh Chandra Satapathy, Chairman, Division V, CSI

Co-chairs

Dr. Nisarg Pathak, KSV, Kadi, Gujarat
Dr. Mukesh Sharma, SFSU, Jaipur

Program Secretary

Er. Sanjay Agal, Pacific University, Udaipur

Members

Mr. Ajay Chaudhary, IIT Roorkee
Dr. Mahipal Singh Deora, BNPG College, Udaipur
Prof. D.A. Parikh, Head, CE, LDCE, Ahmedabad
Dr. Savita Gandhi, Head, CE, Rolwala, G.U., Ahmedabad
Prof. Dr. Jyoti Pareek, Department of Computer Science, Gujarat University
Prof. L.C. Bishnoi, Principal, GPC, Kota
Ms. Bijal Talati, Head, CE, SVIT, Vasad
Er. Kalpana Jain, CTAE, Udaipur
Dr. Harshal Arolkar, Immd. Past Chairman, CSI Ahmedabad Chapter
Mr. Bhavesh Joshi, Advent College, Udaipur
Prof. K.C. Roy, Director, Madhav University, Sirohi
Dr. Pushpendra Singh, Sunrise Group of Institutions
Dr. Sanjay M. Shah, GEC, Gandhinagar
Dr. Chirag S. Thaker, GEC, Bhavnagar, Gujarat
Mrs. Meenakshi Tripathi, MNIT, Jaipur
Mr. Jeril Kuriakose, Manipal University, Jaipur

Chair: Tracks Management

Prof. Vikrant Bhateja, SRMGPC, Lucknow, India

Organising Committee

General Chair

Dr. Dharm Singh, Convener, SIG-WNS, CSI
Dr. Durgesh Kumar Mishra, Chairman, Division IV, CSI

Organising Chair

Dr. Y.C. Bhatt, Chairman, CSI Udaipur Chapter

Organising Co-chairs

Dr. B.R. Ranwah, Immd. Past Chairman, CSI Udaipur Chapter

Local Arrangements Chair

Dr. S.S. Rathore, CTAE, Udaipur

Finance Chair

Prof. S.K. Sharma, Treasurer, CSI Udaipur Chapter

Organising Secretary

Mr. Amit Joshi, CSI Udaipur Chapter

Contents

Resource Management Using Virtual Machine Migrations	1
Pradeep Kumar Tiwari and Sandeep Joshi	
Framework of Compressive Sampling with Its Applications to One- and Two-Dimensional Signals	11
Rachit Patel, Prabhat Thakur and Sapna Katiyar	
Carbon Footprints Estimation of a Novel Environment-Aware Cloud Architecture	21
Neha Solanki and Rajesh Purohit	
Examining Usability of Classes in Collaboration with SPL Feature Model.	31
Geetika Vyas and Amita Sharma	
Do Bad Smells Follow Some Pattern?	39
Anubhuti Garg, Mugdha Gupta, Garvit Bansal, Bharavi Mishra and Vikas Bajpai	
Ultrawideband Antenna with Triple Band-Notched Characteristics	47
Monika Kunwal, Gaurav Bharadwaj, Kiran Aseri and Sunita	
Utilizing NL Text for Generating UML Diagrams	55
Prasanth Yalla and Nakul Sharma	
2-D Photonic Crystal-Based Solar Cell	63
Mehra Rekha, Mahnot Neha and Maheshwary Shikha	
Parallel Implantation of Frequent Itemset Mining Using Inverted Matrix Based on OpenCL	71
Pratipalsinh Zala, Hiren Kotadiya and Sanjay Bhanderi	
Extended Visual Secret Sharing with Cover Images Using Halftoning.	81
Abhishek Mishra and Ashutosh Gupta	

Resonant Cavity-Based Optical Wavelength Demultiplexer Using SOI Photonic Crystals	89
Chandraprabha Charan, Vijay Laxmi Kalyani and Shivam Upadhyay	
A Frequency Reconfigurable Antenna with Six Switchable Modes for Wireless Application	95
Rachana Yadav, Sandeep Yadav and Sunita	
Comparative Analysis of Digital Watermarking Techniques	105
Neha Bansal, Vinay Kumar Deolia, Atul Bansal and Pooja Pathak	
Design and Analysis of 1×6 Power Splitter Based on the Ring Resonator	117
Juhi Sharma	
Performance Evaluation of Vehicular Ad Hoc Network Using SUMO and NS2	127
Prashant Panse, Tarun Shrimali and Meenu Dave	
An Intrusion Detection System for Detecting Denial-of-Service Attack in Cloud Using Artificial Bee Colony.	137
Shalki Sharma, Anshul Gupta and Sanjay Agrawal	
Multi-cavity Photonic Crystal Waveguide-Based Ultra-Compact Pressure Sensor	147
Shivam Upadhyay, Vijay Laxmi Kalyani and Chandraprabha Charan	
Role-Based Access Mechanism/Policy for Enterprise Data in Cloud.	155
Deepshikha Sharma, Rohitash Kumar Banyal and Iti Sharma	
Big Data in Precision Agriculture Through ICT: Rainfall Prediction Using Neural Network Approach.	165
M.R. Bendre, R.C. Thool and V.R. Thool	
Evaluating Interactivity with Respect to Distance and Orientation Variables of GDE Model.	177
Anjana Sharma and Pawanesh Abrol	
Comparative Evaluation of SVD-TR Model for Eye Gaze-Based Systems	189
Deepika Sharma and Pawanesh Abrol	
Identity-Based Key Management	199
Purvi Ramanuj and J.S. Shah	
Firefly Algorithm Hybridized with Flower Pollination Algorithm for Multimodal Functions	207
Shifali Kalra and Sankalp Arora	
Uniqueness in User Behavior While Using the Web	221
Saniya Zahoor, Mangesh Bedekar, Vinod Mane and Varad Vishwarupe	

Prediction of ERP Outcome Measurement and User Satisfaction Using Adaptive Neuro-Fuzzy Inference System and SVM Classifiers Approach	229
Pinky Kumawat, Geet Kalani and Naresh K. Kumawat	
UWB Antenna with Band Rejection for WLAN/WIMAX Band.	239
Monika Kunwal, Gaurav Bharadwaj and Kiran Aseri	
Consequence of PAPR Reduction in OFDM System on Spectrum and Energy Efficiencies Using Modified PTS Algorithm	247
Luv Sharma and Shubhi Jain	
Energy Efficient Resource Provisioning Through Power Stability Algorithm in Cloud Computing	255
Karanbir Singh and Sakshi Kaushal	
Comparative Analysis of Scalability and Energy Efficiency of Ordered Walk Learning Routing Protocol	265
Balram Swami and Ravindar Singh	
A Novel Technique for Voltage Flicker Mitigation Using Dynamic Voltage Restorer	273
Monika Gupta and Aditya Sindhu	
Gaussian Membership Function-Based Speaker Identification Using Score Level Fusion of MFCC and GFCC	283
Gopal, Smriti Srivastava, Saurabh Bhardwaj and Preet Kiran	
Local and Global Color Histogram Feature for Color Content-Based Image Retrieval System	293
Jyoti Narwade and Binod Kumar	
Energy Efficient Pollution Monitoring System Using Deterministic Wireless Sensor Networks	301
Lokesh Agarwal, Gireesh Dixit, A.K. Jain, K.K. Pandey and A. Khare	
Development of Electronic Control System to Automatically Adjust Spray Output.	311
Sachin Wandkar and Yogesh Chandra Bhatt	
Social Impact Theory-Based Node Placement Strategy for Wireless Sensor Networks	319
Kavita Kumari, Shruti Mittal, Rishemjit Kaur, Ritesh Kumar, Inderdeep Kaur Aulakh and Amol P. Bhondekar	
Bang of Social Engineering in Social Networking Sites	333
Shilpi Sharma, J.S. Sodhi and Saksham Gulati	
Internet of Things (IoT): In a Way of Smart World	343
Malay Bhayani, Mehul Patel and Chintan Bhatt	

A Study of Routing Protocols for MANETs	351
Kalpesh A. Popat, Priyanka Sharma and Hardik Molia	
Modelling Social Aspects of E-Agriculture in India for Semantic Web	359
Sasmita Pani, Priyabratta Dash and Jibitesh Mishra	
An Efficient Adaptive Data Hiding Scheme for Image Steganography	371
Sumeet Kaur, Savina Bansal and R.K. Bansal	
A Five-Layer Framework for Organizational Knowledge Management.	381
H.R. Vishwakarma, B.K. Tripathy and D.P. Kothari	
A Survey on Big Data Architectures and Standard Bodies	391
B.N. Supriya, S. Prakash and C.B. Akki	
Generating Data for Testing Community Detection Algorithms	401
Mini Singh Ahuja and Jatinder Singh	
Metamorphic Malware Detection Using LLVM IR and Hidden Markov Model	411
Ginika Mahajan and Raja	
Object-Based Graphical User Authentication Scheme	423
Swaleha Saeed and M. Sarosh Umar	
Efficient Density-Based Clustering Using Automatic Parameter Detection	433
Priyanka Sharma and Yogesh Rathi	
WT-Based Distributed Generation Location Minimizing Transmission Loss Using Mixed Integer Nonlinear Programming in Deregulated Electricity Market	443
Manish Kumar, Ashwani Kumar and K.S. Sandhu	
Use Case-Based Software Change Analysis and Reducing Regression Test Effort.	459
Avinash Gupta and Dharmender Singh Kushwaha	
Tackling Supply Chain Management Through RFID: Opportunities and Challenges	467
Prashant R. Nair and S.P. Anbuudayasankar	
Quality Improvement of Fingerprint Recognition System	477
Chandana, Surendra Yadav and Manish Mathuria	
A Composite Approach to Digital Video Watermarking	487
Shaila Agrawal, Yash Gupta and Aruna Chakraborty	

Effective Congestion Less Dynamic Source Routing for Data Transmission in MANETs	499
Sharmishtha Rajawat, Manoj Kuri, Ajay Chaudhary and Surendra Singh Choudhary	
Implementation and Integration of Cellular/GPS-Based Vehicle Tracking System with Google Maps Using a Web Portal	513
Kush Shah	
Image Segmentation Using Two-Dimensional Renyi Entropy.	521
Baljit Singh Khehra, Arjan Singh, Amar Partap Singh Pharwaha and Parmeet Kaur	
Supervised Learning Paradigm Based on Least Square Support Vector Machine for Contingency Ranking in a Large Power System	531
Bhanu Pratap Soni, Akash Saxena and Vikas Gupta	
Process Flow for Information Visualization in Biological Data.	541
Sreeja Ashok and M.V. Judy	
A Performance Analysis of High-Level MapReduce Query Languages in Big Data	551
Namrata Singh and Sanjay Agrawal	
Variance-Based Clustering for Balanced Clusters in Growing Datasets	559
Divya Saini, Manoj Singh and Iti Sharma	
Conceptual Framework for Knowledge Management in Agriculture Domain	567
Nidhi Malik, Aditi Sharan and Jaya Srivastava	
Modularity-Based Community Detection in Fuzzy Granular Social Networks	577
Nicole Belinda Dillen and Aruna Chakraborty	
Auto-Characterization of Learning Materials: An Adaptive Approach to Personalized Learning Material Recommendation.	587
Jyoti Pareek and Maitri Jhaveri	
Hadoop with Intuitionistic Fuzzy C-Means for Clustering in Big Data.	599
B.K. Tripathy, Dishant Mittal and Deepthi P. Hudedagaddi	
A Framework for Group Decision Support System Using Cloud Database for Broadcasting Earthquake Occurrences.	611
S. Gowri, S. Vigneshwari, R. Sathiyavathi and T.R. Kalai Lakshmi	

Advanced Persistent Threat Model for Testing Industrial Control System Security Mechanisms	617
Mercy Bere-Chitauro, Hippolyte Muyingi, Attlee Gamundani and Shadreck Chitauro	
e-Reader Deployment in Namibia: Fantasy or Reality?	627
Mohammed Shehu and Nobert Jere	
Multi-region Pre-routing in Large Scale Mobile Ad Hoc Networks (MRPR)	637
Majid Ahmad Charoo and Durgesh Kumar Mishra	
Challenges Faced in Deployment of e-Learning Models in India	647
Kamal Kumar Sethi, Praveen Bhanodia, Durgesh Kumar Mishra, Monika Badjatya and Chandra Prakash Gujar	
A Novel Cross-Layer Mechanism for Improving H.264/AVC Video Transmissions Over IEEE 802.11n WLANs	657
Lal Chand Bishnoi and Dharm Singh Jat	
Author Index	669

About the Editors

Dr. Suresh Chandra Satapathy is currently working as Professor and Head, Department of CSE at Anil Neerukonda Institute of Technology and Sciences (ANITS), Andhra Pradesh, India. He obtained his Ph.D. in Computer Science and Engineering from JNTU Hyderabad and M.Tech. in CSE from NIT, Rourkela, Odisha, India. He has 26 years of teaching experience. His research interests are data mining, machine intelligence, and swarm intelligence. He has acted as program chair of many international conferences and edited six volumes of proceedings from Springer LNCS and AISC series. He is currently guiding eight Ph.D. scholars. Dr. Satapathy is also a Sr. Member, IEEE.

Dr. Yogesh Chandra Bhatt graduated from CTAE, Udaipur (1978) and M.Tech. (1980) and Ph.D. (1989) from IIT Kharagpur. Currently, he is serving as Dean (Student Welfare) and Chairman (University Sports Board), Professor and Project In-charge (FIM, FMTC) Department of FMP, CTAE, MPUAT, Udaipur with 36 years of service experience. He is also Professional Agricultural Engineer worked in all divisions of Teaching, Research and Extension wing of the University. He has served as Head of Department for 7 years. He has published more than 50 research papers in international and national journals, seminar and conferences, edited six books, five proceedings and has guided 15 PG students for M.Tech. and Ph.D. degrees. He has completed ten Adhoc research projects. Developed ten prototype technologies on farm mechanization. He has served as Honorary Secretary (2009–2011) and Chairman (2011–2013) of The Institution of Engineers India, Udaipur Local Centre. Organized three international conferences and 18 national conventions and All India Seminar of IEI and ISAE. Published Members Directory of IEI, ULC. He has started Er.M.P. Baya and Mrs. Sheela Baya National Award Rs. 50,000 and Rs. 25,000 and Scholarship for Engineering Students Rs. 60,000 per year from IEI Udaipur. He was Vice Chairman ISAE, Rajasthan Chapter for 8 years and Director of Farm Power and Machinery Group in ISAE for the year 2012–2015 and National Convener of SIG on e-Agriculture in the banner of CSI working on application of ICT techniques in farm machinery. He has received awards of appreciation certificate for outstanding services in MPUAT, Udaipur (2009) and Scroll of Honour from

with Recognition of Eminent Contribution in the field of Agri Engineering (2013) IEI Kolkata. Fellow and Chartered Engineers of IEI, and Life member of ISAE, SESI, CSI and Member ACM. Patron of Rajasthan Agricultural Machinery Association, Jaipur. Working as Chairman of Computer Society of India (CSI) Udaipur Chapter 2015–2016 and Vice Chairman of Association of Computing Machinery (ACM) Udaipur Chapter.

Mr. Amit Joshi has an experience of around 6 years in academic and industry in prestigious organizations of Rajasthan and Gujarat. Currently, he is working as an Assistant Professor in Department of Information Technology at Sabar Institute in Gujarat. He is an active member of ACM, CSI, AMIE, IEEE, IACSIT-Singapore, IDES, ACEEE, NPA and many other professional societies. He is Honorary Secretary of CSI Udaipur Chapter and Honorary Secretary for ACM Udaipur Chapter. He has presented and published more than 40 papers in national and international journals/conferences of IEEE, Springer, and ACM. He has also edited three books on diversified subjects, namely Advances in Open Source Mobile Technologies, ICT for Integrated Rural Development, and ICT for Competitive Strategies. He has also organized more than 25 national and international conferences and workshops including International Conference ETNCC 2011 at Udaipur through IEEE, International Conference ICTCS-2014 at Udaipur through ACM, International Conference ICT4SD 2015—by Springer recently. He has also served on Organising and Program Committee of more than 50 conferences/seminars/workshops throughout the world and presented 6 invited talks in various conferences. For his contribution towards the society, The Institution of Engineers (India), ULC, has given him Appreciation award on the Celebration of Engineers, 2014 and by SIG-WNs Computer Society of India on ACCE, 2012.

Dr. Durgesh Kumar Mishra has received M.Tech. degree in Computer Science from DAVV, Indore in 1994 and Ph.D. degree in Computer Engineering in 2008. Currently, he is working as a Professor (CSE) and Director, Microsoft Innovation Centre at Shri Aurobindo Institute of Technology, Indore, MP, India. He is also a visiting faculty at IIT-Indore, MP, India. He has 24 years of teaching and 10 years of research experience. He completed his Ph.D. under the guidance of late Dr. M. Chandwani on Secure Multi-Party Computation for Preserving Privacy. He has published more than 90 papers in refereed international/national journals and conferences including IEEE, ACM conferences. He has organized many conferences such as WOCN, CONSEG, and CSIBIG in the capacity of conference General Chair and editor of conference proceeding. His publications are listed in DBLP, Citeseer-x, Elsevier, and Scopus. He is a Senior Member of IEEE and has held many positions like Chairman, IEEE MP-Subsection (2011–2012), and Chairman IEEE Computer Society Bombay Chapter (2009–2010). Dr. Mishra has also served the largest technical and profession association of India, the Computer Society of India (CSI) as Chairman, CSI Indore Chapter, State Student Coordinator-Region III MP, Member-Student Research Board, Core Member-CSI IT Excellence Award Committee. Now he is Chairman CSI Division IV Communication at

National Level (2014–2016). Dr. Mishra has delivered his tutorials in IEEE International conferences in India as well as abroad. He is also the programme committee member, and reviewer of several international conferences. He visited and delivered his invited talks in Taiwan, Bangladesh, Singapore, Nepal, USA, UK, and France. He has authored a book on “Database Management Systems.” He had been Chief Editor of Journal of Technology and Engineering Sciences. He has been also serving as a member of Editorial Board of many national and international refereed journals. He has been a consultant to industries and government organizations like Sales Tax and Labour Department of Government of Madhya Pradesh, India. He has been awarded with “Paper Presenter award at International Level” by Computer Society of India. Recently in month of June, he visited MIT Boston and presented his presentation on security and privacy. He has also chaired a panel on “Digital Monozukuri” at “Norbert Winner in twenty-first century” at Boston. Recently, he became the Member of Bureau of Indian Standards (BIS), Government of India for Information Security domain.

Resource Management Using Virtual Machine Migrations

Pradeep Kumar Tiwari and Sandeep Joshi

Abstract Virtualization is a component of cloud computing. Virtualization does the key role on resource (e.g., storage, network, and compute) management and utilized the resource by isolated virtual machines (VMs). A good VMs migration system has an impact on energy efficiency policy. Good resource administration policy monitors the on-demand load management, and also manages the allocation/relocation of the VMs. The key challenges are VMs isolation and migration in heterogeneous physical servers. Threshold mechanism is effective to manage the load among VMs. Previous research shows thrashing is a good option to manage load and migrate the load among VMs during failover and high load. Linear programming (LP), static and dynamic approaches are superior to manage the load through VMs migration. This research proposes threshold-based LP approach to manage load balance and focus on dynamic resource management, load balance goals, and load management challenges.

Keywords Virtual machine · Load balance · Migration · Threshold

1 Introduction

Virtualization technology makes flexible use of resources and provides illusion of high availability of resources to end users as pay-as-use concept [1]. The objectives of effective resource management are optimizing energy consumption, flexible load balance and managing the failover cluster. Research shows an average of 10–15 %

P.K. Tiwari (✉) · Sandeep Joshi
Department of Computer Science & Engineering, Manipal University Jaipur,
Jaipur, India
e-mail: pradeeptiwari.mca@gmail.com

Sandeep Joshi
e-mail: sandeep.joshi@jaipur.manipal.edu

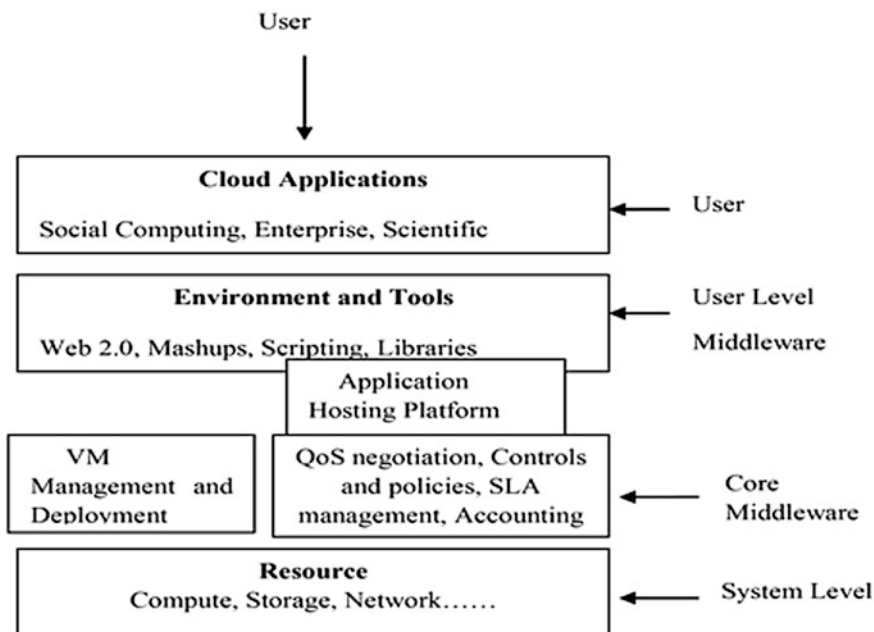


Fig. 1 Model of cloud computing system [3]

capacity is not utilized from 30 % cloud systems. Effective resource utilization can reduce the number of servers [2]. Figure 1 shows layered model of cloud computing system. Cloud computing system can be divided into four parts (User Level, Core Middleware, User Level Middleware, and User Level). Resource management (Compute, Storage, and Network) is a part of system level and it is managed by VM management, which belongs to core middle level. Resource allocation management system can be centralized or decentralized. VM migration system provides isolation, consolidation, and migration of workload. The main purpose of VM migration is to improve performance, fault tolerance, and management of resources [3]. The main focus of cloud system is selection of resources, management of load balance, and maximum utilization of available physical resources via VMs.

Resources are either statically or dynamically assigned to VMs. Static policy is a predefined resource reservation to VM according the need of end user, but in dynamic allocation resources need can be increased and decreased according to the user demands [4].

Server consolidation with migration control is managed by static consolidation, dynamic consolidation, or dynamic consolidation with migration controls.

1.1 Static Consolidation

This approach defines a pre-reserved dedicated resource allocation to VM according the need of end users. VMs resource allocation is based on the total capacity of physical system and the migration does not happen till all demands are changed.

1.2 Dynamic Consolidation

This is a periodicity-based VM migration approach and the migration is based on the current demand. If the required VMs resource demands are higher than the physical available capacity then VMs migrate to another physical server.

1.3 Dynamic Consolidation with Migration Control

This approach gives the stability during high resource demand, hotspot, and frequently changing resources demands. This approach reduces the required number of physical servers and saves the energy consumption. This approach is based on heuristic and round robin mechanism.

VMs migration depends on the resource availability of VMs. If the required load is higher than the thresh value of VM then load will be migrated to another VM. Users’ resource needs, reliability criteria, and scalability terms are clearly mentioned in SLA. Providers provide the resource according to the SLA commitment. To avoid SLA violation, providers provide committed pre-reserved resources and adapt the threshold-based load management policy [5]. Three basic questions arise before VM migration. These questions are: When to migrate? Which VM to migrate? Where to migrate? Figure 2 shows the dynamic resource migration procedure steps. First step monitors the load, if the load is higher than available thresh, then it estimates the load and VMs migrate to another physical machine [1].

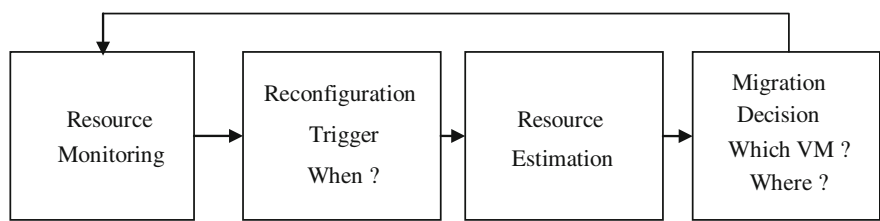


Fig. 2 Dynamic resource management steps [1]

Resource requirements may be different during frequently changing environments, but fixed load does not constrain all time. The best-suited VMs migration strategies are dynamic/heuristic approach for frequently changing environments.

2 VM Migration Management System

Dynamic VM migration management policy manages the on-demand resources availability to users. Upper and lower thresh limits can help to indentify the maximum and the minimum load limits of VMs. Researchers proposed several VM migration techniques to migrate the load from VM to VM and from physical server to physical server. Bin packing approach is good approach for offline resource management but is not effective for the optimal use of CPU. Green computing approach reduces the number of utilized servers and is energy efficient. Memory-aware server consolidation is managed by balloon filter and finger printing system to share VMs locations among heterogeneous servers [6, 7]. Thresh value shows a high capacity of VM. VM load migrates to another VM when a high resources demand arises.

Resources are managed by load monitor and VM allocation/relocation planner. Load monitor monitors the on-demand dynamic load of VMs and compares with available physical resource; relocation planner system plans relocation of VM; and VM controller controls the VM migration among the physical servers and manages the failover cluster. A high demand of resources is the cause of hotspot. Hotspot is detected when server becomes overload due to over demand of resources [8].

LP-heuristic approach [4] proposed linear programming based heuristic WFD (worst fit decreasing), BFD (best fit decreasing), FFD (first fit decreasing), and AWF (almost worst fit decreasing). This is a two-way resource management approach; first policy indentifies VMs and maps the capacity from available physical capacity and second approaches short physical server increasingly according to their capacities with respect to lexicographic order. The objective of linear programming objective is minimization of the required physical server and to map VMs resource availability from hosted physical server.

CP-heuristic [9] resource management system proposed CP (constraint programming) constraint propagation and decision-making research. CP resource allocation results are ten times much faster than integer programming resource scheduling. Heuristic approach uses first fit and best fit methodology for scheduling the short job.

The researcher's visions are green computing and minimal downtime of VM migration. Effective utilization of resources and proper on/off of utilized and non-utilized physical system can contribute to the green computing concept. Researchers continue contributing his/her effort to energy aware resource utilization approach. Energy aware data center location approach [10] proposed MBFD (modified best fit decreasing) and MM (minimization of migration) approach. MBFD optimizes the current VM allocation and chooses the most energy-efficient

nearest physical server for VM to migrate and MM approach minimizes the VM migration needs.

3 Resource Management Goals

Resource management is a semantic relationship between resource availability and resource distribution. Network, compute, and storage are the main resource components. Resource manager manages the resources according to the availability of resources and provides them to the end user on the basis of SLA agreement.

The main goals of resources distribution are (a) performance isolation, (b) resource utilization, and (c) flexible administration.

3.1 Performance Isolation

VMs are isolated from each other and resource utilization of VMs do not affect capacity of another VMs. These VMs are allocated on same physical server. Failure of VM does not affect the performance. Load will migrate to another VM in minimum downtime. Hyper-V provides quick migration and VMware provides the live migration facility. VMs have own reserved network, storage, and compute resources. Resource availability depends on SLA-based user requirements [11, 12].

3.2 Resource Utilization

Resource utilization is based on the maximum consumption of available resources and the minimum energy consumption. Resource manager must give attention to SLA-based on-demand dedicated resource requirements. Resource manager maps the highest requirements on a day, a weekly, and a monthly basis and observes the type of resource needs. This analysis can manage the rush hour resource requirements of end user. VMs resource utilization can be measured by the capacity of physical machine and unused resource can be used as reserved capacity [13].

3.3 Flexible Administration

Resource availability administrator must be able to handle high-load resource and VM migration management in a synchronized manner. VMware uses distributed resource scheduler (DRS) to manage VMs capacity (resource reservations, priorities, and limit) and VM migration. VMware's distributed power management

(DPM) system manages the power on/off management of used/unused VM and plays a vital role in energy aware resource management [14].

Internet small computer system interface (iSCSI) Internet protocol flexibly manages the target storage server and uses storage network protocol (SAN) consolidate storage into storage array. SAN can manage storage consolidation and disaster recovery.

4 Resource Management Challenges

Dynamic resource management with performance isolation, flexible administration, and on-demand resource utilization is not easy to manage simultaneously. Researchers face many problems to utilize resources.

4.1 Flexible VM Migration

Researchers proposed several VM migration methodologies, some are dedicated to resource management and others are energy efficient. Allocation and relocation of VM in heterogeneous environment with load management policy are complicated to manage. Threshold mechanism can map the load of VMs and physical server. Interface management among the VMs and physical servers is complex and tough. Researcher need to more effort to dynamic VMs migration policy [2, 8].

4.2 Storage Management

Researchers proposed [9, 10] virtualized data center management policy. The most popular techniques are pre-copy, post-copy for storage management. CP-heuristic is also a good mechanism to manage data center locations but these techniques are not good enough. Data scattered among heterogeneous locations and gathering of data into a single location in efficient response time is complex.

4.3 Hotspot Management

Hotspot is an overloaded condition of a physical machine. VMs resource access condition is greater than threshold value of the physical machine. This condition is called as hotspot. In static load balance, hotspot condition can be predetermined but in dynamic load, management policy cannot predetermine the hotspot because of its on-demand resource management policy [1].

4.4 SLA Violation

Service level agreement (SLA) defines resource usage patterns of application utilization of storage and computing resource. Cloud service provides pay-as-you-go model. Users only need dedicated pre-reserved on-time resource availability. Cloud service provider gives assurance of quality of service (QoS). Week management of resources, load imbalance, hotspot, and scattered data among heterogeneous servers are main causes of SLA violations [15].

Some service provides offer only guarantee of resource availability rather than performance. The main goal of cloud service providers is the maximum utilization of resources in minimum availability. Server heterogeneity, high resource demand, and failure of cluster affect the resource management and the availability of resources to users. Cloud providers ensure on-demand, robust, scalable, and minimal downtime access services.

4.5 Load Imbalance

End user resource requirement changes dynamically. This can lead to load imbalance in VMs. VMs use physical machine resource capacity, sometimes VMs load may be higher to a physical machine, then VM will migrate to another physical machine. Some physical machines are highly loaded and some are less loaded. It may cause discrepancy in utilizations of physical servers. Overloaded VMs downtime is higher than low-loaded ones [1].

5 Load Balance Concept and Management

High performance computing (HPC) can be achieved by DRS in minimal mean-time. Threshold-based load migration with LP is a superior approach to load migration. This approach is static during the distribution of VMs ID allocation and dynamic in load migration among the VMs. This article proposes two algorithms: first algorithm allocates ID to all high-load, low-load, and average load VMs and second algorithm controls the process of load migration among VMs.

First algorithm counts and locates the VMs ID of single physical system. It classifies the VMs according to high, average, and low load. Its first array stores the ID of high-load VMs and second array stores the low-load VMs information. Low load and high load are measured by average thresh value (Ath). Average thresh value is a maximum queue length of an individual VM. VMs can be distributed according to high-load, low-load, and remaining (average load) VMs store in average array.

Algorithm 1: Virtual machine Identification

```

1: V - Total number of VMs in single host
2: For each VM [ID] -  $VM \in V$  // Assign unique ID to VMs
3: VM [ID] load == Ath (Thresh value of VM) //Check
the current load from available thrash
    if VM [ID] > Ath
        Send VM [ID] to Hth [array] //Store in high
load array
    else if VM [ID] < Ath
        Send VM [ID] to Lth [array] // Store in low
load array
    else
go to Ath array // store in maximum trash array
4. end for

```

Second algorithm specifies the load distribution form high-load to low-load VM. Algorithm can find the high-load VM from high [array] and low-load VM from low [array]. It checks for the transfer condition, if transfer condition is true then the load is transferred from high load to low load. If low-load VM queue length is full then high-load VM load will be transferred to another low-load VM. High-load VM load will come equal to Atv then high-load VM will be transferred to average array exit, if high-load VM current load is low then the VM moves to low array and then exits.

Algorithm 2: Load Balance Management

```

1: Find the VM form Hth > Ath
2: Find the VM from Lth
3: [ok = Lth [VM] < Ath < Hth [VM] /* lock the both
(Hth[VM] and Lth[VM], Check the condition and
continue transfer the load */
    if (current Lth(VM) == Ath) // during the
load transfer
        Then Lth[VM] move to Ath array
        go to step 2 // take the VM form Lth array
        repeat step 3 //
if Hth [VM] == Ath
    Move current Hth [VM] to Ath
else
    go to Lth array
4: repeat step 1 to 3
5: end

```

Algorithm manages low-load and high-load VMs. Meantime of response will be shorter and effective. Load balance must be flexible, performance-based resources utilization. Thresh valve of VM defines the high capacity of VM and does the help to find low-load and high-load VMs.

6 Conclusion

Traditional computing systems use cluster and grid computing based resource management mechanism. Cloud service provides virtualized distributed resource management policy. Good resource management policies maximize the resource utilization. Resource management policy must be able to scale the available resource and users demands. In this paper we have discussed resource management goals, VM migration policies, challenges of resource management, and load balance concept and management. This paper has proposed an effective load balance algorithm. This algorithm has proposed a way to find overloaded VMs and underloaded VMs with load transfer mechanism. Researchers have done lots of work in load balance, VM migration, and server consolidation with migration policy, but still there is a need for intelligent live migration mechanism.

References

1. Mishra, Mayank, Anwesha Das, Purushottam Kulkarni, and Anirudha Sahoo. "Dynamic resource management using virtual machine migrations." *Communications Magazine*, IEEE 50, no. 9 (2012): 34–40.
2. Ahmad, Raja Wasim, Abdullah Gani, Siti Hafizah Ab Hamid, Muhammad Shiraz, Abdullah Yousafzai, and Feng Xia. "A survey on virtual machine migration and server consolidation frameworks for cloud data centers." *Journal of Network and Computer Applications* 52 (2015): 11–25.
3. Hussain, Hameed, Saif Ur Rehman Malik, Abdul Hameed, Samee Ullah Khan, Gage Bickler, Nasro Min-Allah, Muhammad Bilal Qureshi et al. "A survey on resource allocation in high performance distributed computing systems." *Parallel Computing* 39, no. 11 (2013): 709–736.
4. Ferreto, Tiago C., Marco AS Netto, Rodrigo N. Calheiros, and César AF De Rose. "Server consolidation with migration control for virtualized data centers." *Future Generation Computer Systems* 27, no. 8 (2011): 1027–1034.
5. Beloglazov, Anton, and Rajkumar Buyya. "Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers." In *Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science*, p. 4. ACM, 2010.
6. Hu, Liang, Jia Zhao, Gaochao Xu, Yan Ding, and Jianfeng Chu. "HMDC: Live virtual machine migration based on hybrid memory copy and delta compression." *Appl. Math* 7, no. 2L (2013): 639–646.
7. Medina, Violeta, and Juan Manuel García. "A survey of migration mechanisms of virtual machines." *ACM Computing Surveys (CSUR)* 46, no. 3 (2014): 30.
8. Beloglazov, Anton, and Rajkumar Buyya. "Energy efficient resource management in virtualized cloud data centers." In *Proceedings of the 2010 10th IEEE/ACM International*

- Conference on Cluster, Cloud and Grid Computing, pp. 826-831. IEEE Computer Society, 2010.
9. Zhang, Yuan, Xiaoming Fu, and K. K. Ramakrishnan. "Fine-Grained Multi-Resource Scheduling in Cloud Datacenters." In The 20th IEEE International Workshop on Local and Metropolitan Area Networks (LANMAN 2014). 2014.
 10. Beloglazov, Anton, Jemal Abawajy, and Rajkumar Buyya. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." *Future generation computer systems* 28, no. 5 (2012): 755–768.
 11. Gupta, Diwaker, Ludmila Cherkasova, Rob Gardner, and Amin Vahdat. "Enforcing performance isolation across virtual machines in Xen." In *Middleware 2006*, pp. 342–362. Springer Berlin Heidelberg, 2006.
 12. Nathan, Senthil, Purushottam Kulkarni, and Umesh Bellur. "Resource availability based performance benchmarking of virtual machine migrations." In *Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering*, pp. 387–398. ACM, 2013.
 13. Isci, Canturk, Jiuxing Liu, Bülent Abali, Jeffrey O. Kephart, and Jack Kouloheris. "Improving server utilization using fast virtual machine migration." *IBM Journal of Research and Development* 55, no. 6 (2011): 4–1.
 14. https://pubs.vmware.com/vsphere-50/index.jsp#com.vmware.vsphere.vm_admin.doc_50/GUID-E19DA34B-B227-44EE-B1AB-46B826459442.html, 2015.
 15. Garg, Saurabh Kumar, Adel Nadjaran Toosi, Srinivasa K. Gopalaiyengar, and Rajkumar Buyya. "SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter." *Journal of Network and Computer Applications* 45 (2014): 108–120.
 16. https://pubs.vmware.com/vsphere-50/index.jsp#com.vmware.vsphere.vm_admin.doc_50/GUID-E19DA34B-B227-44EE-B1AB-46B826459442.html, 2015.
 17. Li, Kangkang, Huanyang Zheng, Jie Wu, and Xiaojiang Du. "Virtual machine placement in cloud systems through migration process." *International Journal of Parallel, Emergent and Distributed Systems* ahead-of-print (2014): 1–18.

Framework of Compressive Sampling with Its Applications to One- and Two-Dimensional Signals

Rachit Patel, Prabhat Thakur and Sapna Katiyar

Abstract Compressive sampling emerged as a very useful random protocol and has become an active research area for almost a decade. Compressive sampling allows us to sample a signal below Shannon Nyquist rate and assures its successful reconstruction with some limitations on signal, that is, signal should be sparse in some domain. In this paper, we have used compressive sampling for an arbitrary one-dimensional signal and two-dimensional image signal compression and successfully reconstructed them by solving L1-norm optimization problems. We also have showed that compressive sampling can be implemented if a signal is sparse and incoherent through simulations. Further, we have analyzed the effect of noise on the recovery.

Keywords Basis function • Compressive sampling • Incoherent signal • L1-norm • Sparse signal

1 Introduction

Today we are moving towards digital domains, but origination of the signal most of the times be analog. Therefore, analog-to-digital converting systems are required but these systems bounded by criteria that sampling frequency should be greater than twice of the analog signal frequency (Shannon Nyquist criteria) [1]. But if frequency of signal is very high then it is very tiresome to use Nyquist criteria

Rachit Patel (✉) • Sapna Katiyar
ABES Institute of Technology, Ghaziabad, UP, India
e-mail: rachit05081gece@gmail.com

Sapna Katiyar
e-mail: sapna_katiyar@yahoo.com

Prabhat Thakur
Jaypee University of Information Technology, Solan, HP, India
e-mail: thakurprabhat2010@gmail.com

because the number of samples will be very large. It becomes costly and sometimes infeasible to store and process such large number of samples.

Nevertheless, if somehow we can overcome Shannon Nyquist criteria that we can reconstruct original signal by using a very less number of samples as compare to Nyquist criteria, then problem of storage and processing of large data can be solved. This problem may be solved by compressive sampling [2–4] a random approach if signal is sparse in some domain.

Compressive sampling uses a very less number of samples as compared to Shannon Nyquist rate which reduces the hardware and software loads and then signal is recovered by using various recovery mechanisms [5–7]. Compressive sampling uses a random matrix to form out linear random projections of signals with most of the desired information. It is possible due to two properties of signal, i.e., sparsity and incoherence [8]. Sparsity refers to the property of signal according to which information present in signal is very less as compared to the bandwidth occupied by the signal. Incoherence is a property of sparse signal to get transformed into desired domain. Desired domain is the domain in which signal is sparse. If signal is more sparse, i.e., low sparsity level, its reconstruction will be better as compared to less sparse signal. Basically, incoherence refers to not coherent, i.e., the dictionary (domain) elements should be independent to the sampling matrix.

The paper is organized as follows. Section 2 provides some background on compressive sampling, mathematical model, and signal reconstruction by solving optimization problems. In Sect. 3 we use compressive sampling for one-dimensional and two-dimensional signals compression and its successful reconstruction. Section 4 presents simulation results on its performance and Sect. 5 presents conclusion and future scope.

2 Background

2.1 Compressive Sampling

Compressive sampling (CS) is an emerging theory which allows us to project random measurements of signal of interest so that we can sample the signal at information rate and not at its ambient data rate. This reduces the number of samples to represent a signal. These less number of samples can be stored easily and processing of such small number of samples can be performed efficiently. But to apply compressive sampling on the signal, signal should be sparse and incoherent.

Sparse Signal: For a signal to be sparse, only some of the components should have considerable magnitude and all other components should have very less magnitude, i.e., closer to zero.

Incoherent Signal: For two signals to be coherent, they should be independent of each other.

2.2 Mathematical Approach for Compression

Consider a signal r which is sparse. r is said to be sparse if it can be represented as a linear combination of basic functions where some of the coefficient's magnitude are significant and all others have zero magnitude

$$r = \psi c$$

ψ —basis functions and c —basis coefficients

For compressive sampling, a random matrix $\emptyset$ needs to project random projections or random measurements

$b = \emptyset r$, here b is random measurement vector

2.3 Reconstruction

Now we need to reconstruct back r from b

$$b = \emptyset \psi c \quad (1)$$

By solving equation we can find basis coefficient c .

Information of c leads us towards recovery solution

$$r = \psi c$$

2.4 Optimization Problem Formulation

The equation we have to solve, i.e., (1) is an underdetermined system as number of equations is less than number of unknowns. So we need to use norm minimization techniques to solve above problem.

Mathematically norm provides the total size or positive lengths of all vectors in a vector space or matrices. Generally, Norm n of vectors x is defined as

$$\|x\|_n = \sqrt[n]{\sum_i \|x\|^n} \text{ Where, } n \in R$$

Frequently using norms are L0, L1, L2 but here we use L1-norm.

L1-norm: L1-norm is defined as $\|x\|_1 = \sum_i |x|$

L1 optimization problem is formulated as

$$\min \|x\|_1 \quad \text{subject to } |Ax = b|$$

Above problem can be solved using least square optimization

$$x = A^+ b,$$

where A^+ –Pseudoinverse of A

Even though this method is easy to compute it is not necessary that it provides best solution. That is why we use L1-norm optimization.

So our optimization problem can be formulated as

$$\min \|c\|_1 \quad \text{subject to } |(\mathcal{O}\psi)c = b|$$

2.5 L1 Optimization Solution

L1 optimization problems can be solved by using linear or nonlinear programming algorithms such as greedy-type orthogonal matching pursuit, basic pursuit [9].

3 Applications in One- and Two-Dimensional Signal

Reduced load on hardware and software leads us to use compressive sampling in all possible fields such as compression, image compression, speech compression, audio and video compression, wireless sensor networks, etc. But here we apply compressive sampling on one-dimensional and two-dimensional image signals.

3.1 One-Dimensional Signal Compression and Recovery

In our daily life, number of times we deal with one-dimensional signal such as audio signals speech signals and we need to sample these signal for performing some digital operation on these signal. Less number of samples can be processed easily with a short processing time. So we go for compressive sampling of such signals if they are sparse. Complete recovery of signal depends on sparsity level (SL) and compression ratio (CR). Sparsity level is a number of components having significant magnitude. Compression ratio is the ratio that up to what level we have compressed the signal, e.g., $N/10$, where N is the total number of samples present in the signal.

If sparsity level is low, recovery will be better. If compression ratio is more, recovery will be better. Consider one-dimensional signal r having length n .

r_{n*1} can be represented with the help of basic functions and its coefficients

$$r_{n*1} = \psi_{n*n} c_{n*1} \quad (2)$$

$\psi_{n*n} - n * n$ Matrix of basis function

$c_{n*1} - n * 1$ Vector of basis coefficients

For random measurements after random sampling we use measurement matrix $\varnothing_{m*n}$

$$b_{n*1} = \varnothing_{m*n} r_{n*1}$$

$\varnothing_{m*n}$ —Measurement Matrix.

$$b_{n*1} = (\varnothing_{m*n} \psi_{n*n}) c_{n*1} \quad \text{using (2)}$$

Above equation needs to be solved using L1-norm optimization. L1-norm optimization problem is formulated as

$$\min \|c_{n*1}\| \quad \text{subject to } (\varnothing_{m*n} \psi_{n*n}) c_{n*1} = b_{n*1} \quad (3)$$

Reconstruction using above solution

$$\widehat{r_{n*1}} = \psi_{n*n} c_{n*1}$$

3.2 Recovery Error (Rerr)

Recovery error is a very important parameter that gives us the error for successful recovery of the signal and is defined as

$$\text{Rerr} = \text{norm2}(r - \hat{r})$$

3.3 Two-Dimensional Signal Compression and Recovery

Two-dimensional signals like image can also be compressed using its Fourier or wavelet domain where image shows some sparse nature. A mathematical approach for image remains same as for signals but we choose basis functions either on Fourier or wavelet domain. Instead of a one-dimensional vector we deal with a two-dimensional matrix.

3.4 Effect of Noise on Recovery Error

Noise is an undesired signal that may affect the performance of the system. So we analyzed the effects of noise on our system of compressive sampling, i.e., how recovery error is going to vary with respect to noise.

For compressive sampling, noise may affect the sampled values and mathematical equation for sampled signal will be defined using Eq. (2)

$$b_{m*1} = \Phi_{m*n} r_{n*1} + n_{m*1}$$

where n_{m*1} —noise vector

Recovery procedure will be same as in Eq. (3), i.e., we need to solve L1-norm minimization problem

$$\min \|c_{n*1}\|_1$$

$$\text{such that } \|b_{m*1} - (\Phi_{m*n} \psi_{n*n})c_{n*1}\|_2 \leq \varepsilon$$

But due to addition of noise, values of vector b change. Thus, above problem can be solved with affected value of b . So value of coefficient vector c also varies from the desired values. By this way reconstruction or recovery gets affected.

4 Simulations and Results

For implementation of all algorithms, we used Matrix Laboratory (Matlab) on a standard computer. The L1-magic toolbox is used to achieve the solution of L1-norm optimization problems.

4.1 Signal Reconstruction

We considered a signal in time domain and made it sparse in frequency domain by taking all frequency domain coefficients zero which are below some threshold value. Here threshold value is assumed to be one-fifth of the maximum amplitude of the coefficients. We used same procedure used for signal compression and recovery of original signal in Sect. 3.1. We sampled the signal using sampling rate which is ten times lesser than Nyquist rate and successfully reconstructed the original signal as shown in Fig. 1. Here we varied the sparsity level of signal and analyzed its results on recovery error. In addition to this, we also analyzed the effect of variation of compression ratio on recovery error.

Fig. 1 Original signal versus recovered signal

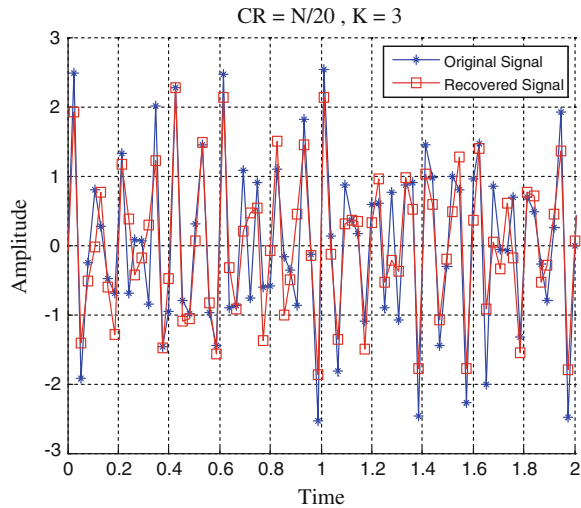


Fig. 2 Recovery error versus number of samples after compression with different sparsity level

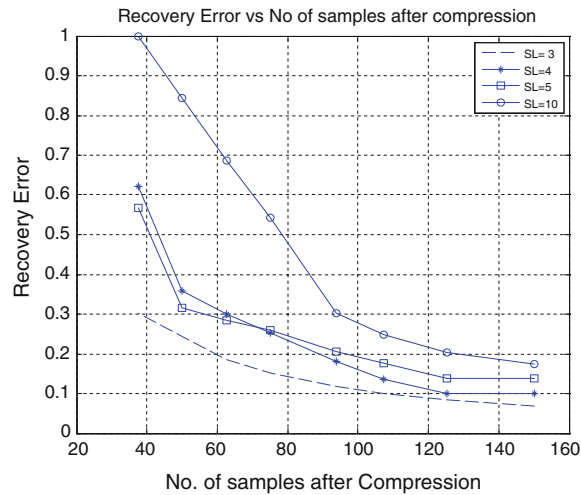
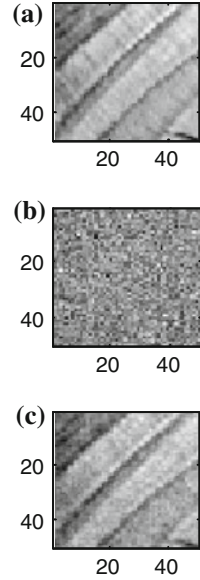


Figure 2 explains the behavior of recovery error with variation in compression ratio for different sparsity levels. It is clear that with the increase in compression ratio, recovery error increases. More the sparsity level is more the recovery error. So it confirms that the compressive sampling can be implemented on the signals having sparse nature otherwise recovery or reconstruction cannot be done successfully. Recovery error relies on sparsity level and compression ratio. If signal is sparser, it can be compressed more and can be reconstructed successfully.

Fig. 3 Original image versus recovered image. **a** Original signal. **b** Recovered image using least square method. **c** Recovered image using basic pursuit



4.2 Two-Dimensional Signal's Reconstruction

Further, we compress the image by taking random measurements and recover the image using least square method (LSM) and basis pursuit (BP, an algorithm to solve L1-norm optimization problem).

Figure 3 contains three images, namely original image, recovered image using least square method, and recovered image using basis pursuit. From here it is clear that image recovered using least square method is much distorted but image recovered using basic pursuit is almost similar to original image.

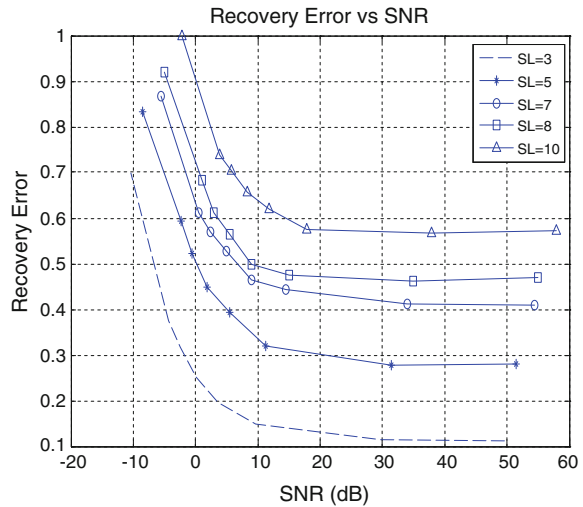
4.3 Effect of Noise on Recovery Error

We have seen in Sect. 2.4 how noise affects our recovery error. Here we analyzed the effect of noise on the recovery error for different values of sparsity level (Fig. 4).

5 Conclusion and Future Scope

Compressive sampling appears to be a revolutionary technique for data acquisition and successful reconstruction. We have implemented this technique for one-dimensional signal as well as two-dimensional image signal and successfully

Fig. 4 Effect of SNR on recovery error



recover them from compressive random measurements. We have analyzed recovery error due to variations in sparsity level and compression ratio and assured that successful reconstruction of signal relies on sparsity level and compression ratio. Effect of noise is also considered in compressive sampling and we verified that with increase in SNR, recovery error decreases, that is, according to our system expectations.

In future scope, we can use this technique for compression of another kind of one-dimensional or multidimensional signal if they are sparse in some domain. This technique can play important role in microwave applications where sampling rate is very high due to high frequencies.

References

1. Haykin, S.: Communication Systems, 4th ed., John Wiley and Sons, NewYork, 2001.
2. Donoho, D.: Compressed sensing, IEEE Trans. Inform. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
3. Candès, E., Romberg, J., and Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489–509, 2006.
4. Candès, E. J.: Compressive sampling, in Proceedings of the International Congress of Mathematicians, vol. 3, pp. 1433–1452, 2006.
5. Tropp, J. and Gilbert, A.C.: Signal recovery from partial information via orthogonal matching pursuit, IEEE Trans. Inform. Theory, vol. 53, no. 12, pp. 4655–4666, 2007.
6. Candès, E. and Tao, T.: Decoding by linear programming, IEEE Trans. Inform. Theory, vol 51, no 12, pp. 4203–4215, December 2005.
7. Chen, S. S., Donoho, D. L. and Saunders, M. A.: Atomic decomposition by basis pursuit, SIAM J. Sci Comp., 20(1):33–61, 1999.

8. Candès, E. and Romberg, J.: Sparsity and incoherence in compressive sampling, *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.
9. Cai, T. T., and Wang, L.: Orthogonal matching pursuit for sparse signal recovery with noise, *IEEE Trans. on inform. Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.

Carbon Footprints Estimation of a Novel Environment-Aware Cloud Architecture

Neha Solanki and Rajesh Purohit

Abstract Carbon footprints are increasing with a huge rate and the IT world is also contributing in this increase. In cloud computing, with the growth of demand for high performance computing infrastructure, number of data centers has increased. To cater the demand of high availability, the data centers are kept running round the clock. This causes high energy consumption and eventually increases in carbon footprints, which is harmful for environment. In addition to this, high energy consumption leads to costlier business. In this paper, a novel architecture for cloud is proposed by introducing an energy-aware service provider layer. The responsibility of this layer is to monitor and control the performance of cloud data centers for reducing energy consumption and carbon footprints. Live migration of virtual machines among physical machines is applied as basic technique for reducing the energy consumption.

Keywords Bin-packing algorithms • Carbon footprints • Consolidation • Data centers • Live migration

1 Introduction

Cloud computing delivers on-demand computing resources over the Internet on the basis of pay-as-you-go model. According to National Institute of Standards and Technology, USA, “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction” [1]. It provides three types of services:

Neha Solanki (✉) • Rajesh Purohit
Computer Science and Engineering Department, M.B.M Engineering College,
Jai Narain Vyas University, Jodhpur, India
e-mail: nehasolanki2789@gmail.com

Rajesh Purohit
e-mail: rajeshpurohit@jnvu.edu.in

software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).

Virtualization is used in cloud computing which shares the underlying hardware infrastructure and through virtual machines it provides the computing resources. It gives full control to cloud provider's administrator for virtual machine allocation, which results in efficient utilization of the resources.

Cloud services are run by data centers. A single data center may consist of a large number of physical machines. These data centers consume huge amount of electricity which increases carbon footprints and the computational cost. According to National Resource Defense Council's report, energy consumed by data centers in US in 2013 was estimated as 91 billion kWh of electricity which is enough to power all the houses in New York City twice and this is annual output of 34 large (500-mW) coal-fired power plants [2]. Data center's electricity consumption is supposed to increase to around 140 billion KWh annually by 2020, which is equal to 50 power plants' annual output and emitting nearly 150 million metric tons of carbon pollution annually [2]. Hence from 2013 to 2020, data center energy consumption is projected to increase roughly 53 %.

It is necessary to reduce this high energy consumption so that carbon footprints also get reduced. The main objective of this paper is to estimate the carbon footprints of the proposed architecture for environment-aware cloud computing. CloudSim is a simulation framework developed at University of Melbourne which allows modeling/simulation and experimentation of proposed solution on specific system design issues for investigation at abstract level [3]. To convert the results of estimated energy consumption to carbon footprints, defra conversion factor is applied [4].

2 Related Work

In recent years, many researchers have invested their time on energy-saving architecture and techniques of cloud. Verma et al. [5] have developed the architecture of pMapper, a power-aware application placement controller with heterogeneous virtualized server clusters. It minimizes power and migration cost, while fulfilling the performance requirement.

Beloglazov et al. [6] developed an architectural framework and principles for energy-efficient cloud computing. This architecture aimed to improve energy efficiency of the data center, while delivering the negotiated quality of service (QoS). For this, live migration concept is used to save energy of cloud data centers. Aboozar et al. [7] have proposed decision support-as-a-service (DSaaS) architecture which is divided into two subsystems: cloud side and user side to help managers for making fast and precise decisions for energy saving. Yanfeiet et al. [8] have proposed system architecture of EECLOUD, in which the data center is divided into several time-sub-clusters to which jobs with similar runtime are assigned. It uses live migration for saving energy consumption. Jinhai et al. [9] have designed

energy-efficient architecture of cloud data center which is divided into two types of controller units: global controller and local controller. It also uses live migration to reduce energy consumption.

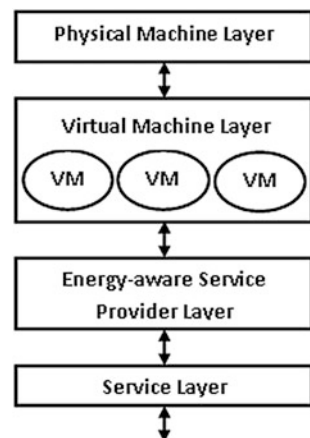
3 Environment-Aware Architectural Framework of Cloud Data Centers

The abstract view of proposed environment-aware architecture is divided into four layers, i.e., service layer, energy-aware service provider layer, virtual machine layer, and physical layer as depicted by Fig. 1.

1. Service Layer: This layer provides client interface for submitting their request through web to enable time, location, and device independency.
2. Energy-aware Service Provider Layer: This layer acts as an interface between users and cloud infrastructure and it is responsible for conserving energy of cloud data centers. It uses various live migration strategies to consolidate virtual machine to save energy [10].
3. Virtual Machine Layer: This layer consists of virtual machines (VM) which can dynamically start and stop physical machines according to incoming users' requests. On a single physical machine, multiple virtual machines can run applications concurrently.
4. Physical Machine Layer: This layer consists of computing servers which provide the hardware infrastructure like computing resource, storage resource, etc. This hardware infrastructure is used for making virtualized resources to fulfill users' service demands.

The proposed environment-aware architecture shown in Fig. 2 reduces energy consumption of cloud data centers by consolidating the allocated virtual machines

Fig. 1 Abstract view of environment-aware architecture of cloud



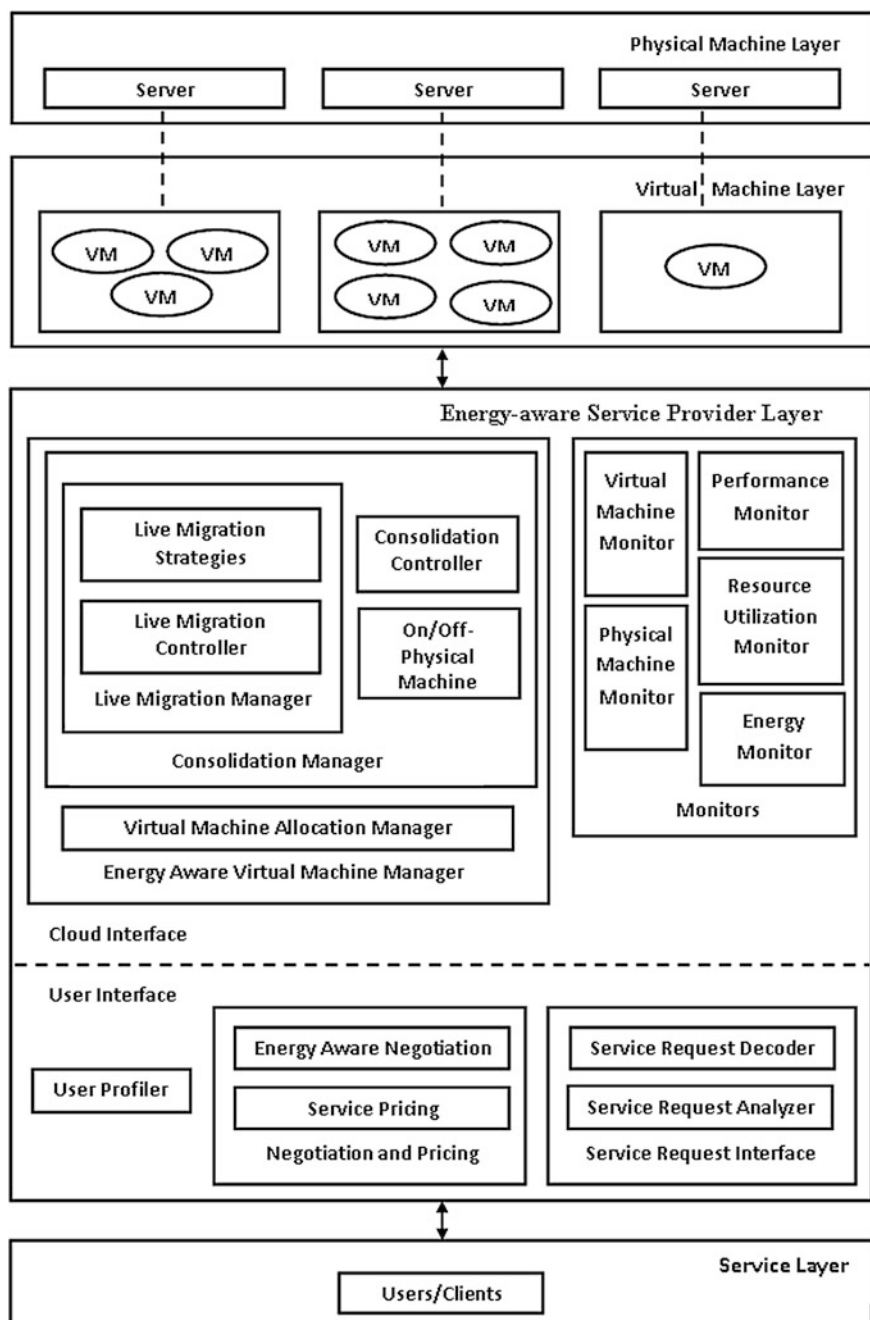


Fig. 2 Environment-aware architecture of cloud

to minimize the number of currently running physical machines and switching off the idle physical machines. To fulfill the requirement of new virtual machines, these switched-off physical machines are switch-on.

In the *Service Layer*, users/clients submit their service request to the cloud. *Energy-aware Service Provider Layer* which acts as an interface between users and cloud infrastructure is divided into two sub layers: *User Interface* and *Cloud Interface*.

User Interface sub layer consists of: *User Profiler*, *Negotiation and Pricing*, and *Service Request Interface* module. *User Profiler* collects information about the users like their priorities, interest in specific services, etc., so that some special advantages can be given to the users. *Negotiation and Pricing* is subdivided into two parts based on the distinction of their functionality, i.e.,: *Energy-aware Negotiation* and *Service Pricing*. *Energy-aware Negotiation* is responsible for negotiating with the users according to the user's QoS demand and energy-aware technique to settle the service level agreement (SLA) and penalties with specified prices between the cloud provider and user. *Service Pricing*, as the name depicts, deals with pricing issues with users according to the type and scale of service opted. The third module of *User Interface* sub layer, i.e., *Service Request Interface* is subdivided into *Service Request Decoder* and *Service Request Analyzer*. When the service is requested, first it is identified by the *Service Request Decoder* and sent to *Service Request Analyzer* to decide whether the requested service can be granted or not according to the SLA and availability of resource.

Cloud Interface sub layer consists of *Service Request Scheduler*, *Monitors*, and *Energy-aware Virtual Machine Manager* Module. Gathered user's service requests are scheduled according to predefined policy through *Service Request Scheduler* [11]. *Virtual Machine Monitor*, *Physical Machine Monitor*, *Performance Monitor*, *Resource Utilization Monitor*, and *Energy Monitor* are the parts of *Monitors* module which provides its monitored information to the other components of the cloud interface so that they can perform their tasks.

Virtual Machine Monitor and *Physical Machine Monitor* are responsible for probing into system and keeping the status of virtual machines and physical machines, respectively, by counting and identifying which virtual machines and physical machines are on/off. Performance of the user's service request according to the SLA is monitored by the *Performance Monitor*. *Resource Utilization Monitor* interacts with the virtual machines to monitor the amount of resources utilized while processing the service requests. Energy consumption by each physical machine is monitored by *Energy Monitor*.

The third module of *Cloud Interface* sub layer is *Energy-aware Virtual Machine Manager*, which is responsible for saving energy and carbon footprints by allocating new virtual machine request on physical machine consolidating the allocated virtual machine to minimize the current running physical machine and switching off the idle physical machines. It is carried out in two phases by *Virtual Machine Allocation Manager* and *Consolidation Manager*. *Virtual Machine Manager* deals with allocating new request of virtual machine on physical machine. This process of allocation is modeled through classical bin-packing algorithm, e.g., best-fit

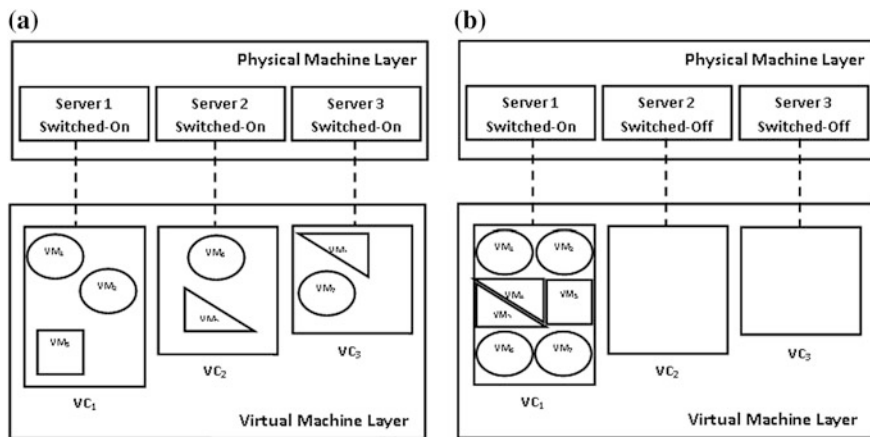


Fig. 3 Consolidation of virtual machines in environment-aware architecture of cloud. **a** Before consolidation. **b** After consolidation

algorithm, first-fit algorithm, best-fit decreasing algorithm, etc. The bin-packing algorithm is analogous to the allocation process, in which packets of given various sizes (to model Virtual Machines) are attempted to be packed into a minimum number of containers/bins (to model physical machines) [12]. The work of *Consolidation Manager* is to consolidate the allocated virtual machine by live migrating them among other available physical machines and switching off the idle physical machine(s) as shown in Fig. 3. Virtualization capacity is the maximum number of virtual machine which can be allocated on an individual physical machine depending on the resource availability of individual physical machine. It is denoted by VC_i and represented by rectangular area for individual physical machines. Virtual machine is denoted by VM_j and is represented by circle, triangle, and square shapes and sizes according to the resource requirement to show heterogeneity.

Consolidation Manager is composed of *Live Migration Manager*, *Consolidation Controller*, and *On/Off-Physical Machine Controller*. *Live Migration Manager* deals with migrating virtual machine from underloaded or overloaded physical machine. Migration from underloaded physical machine is carried out to save energy while migration from overloaded physical machine is carried out to avoid violation of any QoS. *Live Migration Controller* is responsible for enabling the process of live migration by using various *Live Migration Strategies*. *Consolidation Controller* is responsible for initiating and terminating consolidation, and its output is used by *On/Off-Physical Machine Controller* to switch off the idle physical machine to save energy and carbon footprints, and to switch-on physical machine for fulfilling new virtual machine requirement.

Physical Machine Layer consists of multiple servers on which various service requests get executed. Virtualization hides the infrastructure complexity of underlying hardware and abstracts the physical infrastructure. It allows creating multiple

virtual machines on single server to improve utilization of server. *Virtual Machine Layers* consist of multiple virtual machines; they are operating system independent and can run multiple applications concurrently on single server to properly utilize the hardware resources.

4 Methodology

The whole method of consolidating virtual machines is carried out in two phases: allocation and optimization. In the first phase, newly requested VMs are placed into suitable physical machines on the basis of CPU utilization using bin-packing algorithms. In this work, two bin-packing algorithms are used: first-fit decreasing and first-fit. In first-fit decreasing algorithm, the requested virtual machines are sorted in nonincreasing order according to their CPU utilization and the suitability of physical machine is checked. The virtual machine is allocated to the first physical machine, which satisfies the criteria. In the first-fit algorithm, similar steps are carried out except sorting of the requested virtual machine. In the second phase, currently allocated VMs are optimized by migrating VMs among physical machines. Migration of VMs depends upon the predefined lower utilization threshold and upper utilization threshold of physical machines. Minimum migration time (MMT) policy is used to select VM for migration [6]. This policy selects that VM which takes minimum time for migration. After selecting the VM for migration, again bin-packing algorithms are used to place them into suitable physical machines.

5 Experiments and Results

The proposed architecture has been evaluated by simulation using CloudSim toolkit. Data center consists of physical machines. Users' requests are submitted through VMs. Two types of physical machines are used consisting of 2660 and 1860 MIPS and both having two processing elements. The lower and upper utilization thresholds are set to 40 and 90 %, respectively. In this experiment, first-fit decreasing and first-fit algorithms with minimum migration time policy are used. For comparative purpose, simulation of non environment-aware policy is also carried out, which does not consolidate VMs, i.e., it does not have the migration of VMs.

Table 1 depicts the simulation result for energy consumption, carbon footprints, SLA violation, and number of migrations. Nonenvironment-aware policy has very high energy consumption in comparison to minimum migration time policy with first-fit decreasing and first-fit algorithm.

The comparison of carbon footprints between non environment-aware policy and minimum migration time policy with first-fit decreasing algorithm is shown in Fig. 4, and between non environment-aware policy and minimum migration time

Table 1 Results

Policy	Energy consumption (kWh)	Carbon footprints (kg CO ₂ e)	SLA violation (%)	No. of migration
Non environment-aware	23.33	12.54		
MMT-first-fit	3.90	2.10	0.32	1375
MMT-first-fit decreasing	3.56	1.91	2.8	1013

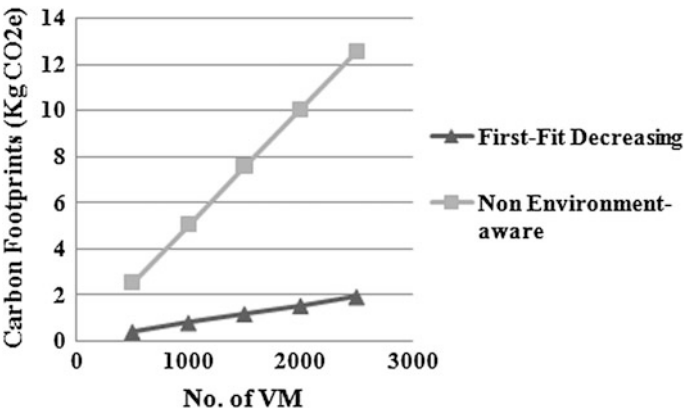


Fig. 4 Carbon footprints: first-fit decreasing and non environment-aware

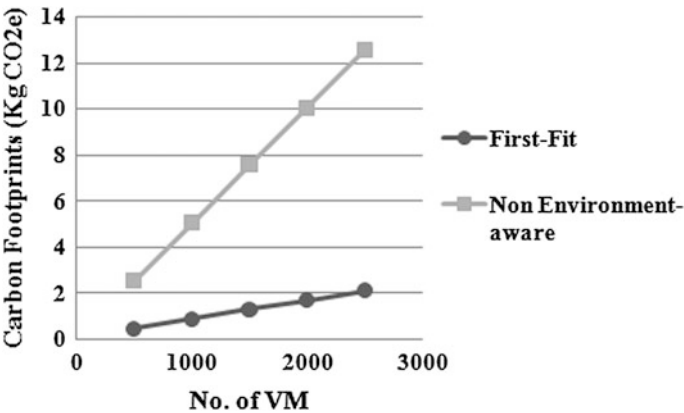


Fig. 5 Carbon footprints: first-fit and non environment-aware

policy with first-fit algorithm is shown in Fig. 5. They show that with migration of VMs carbon emission can be saved, both first-fit decreasing and first-fit algorithms using minimum migration time policy have less carbon footprints in comparison to non environment-aware policy.

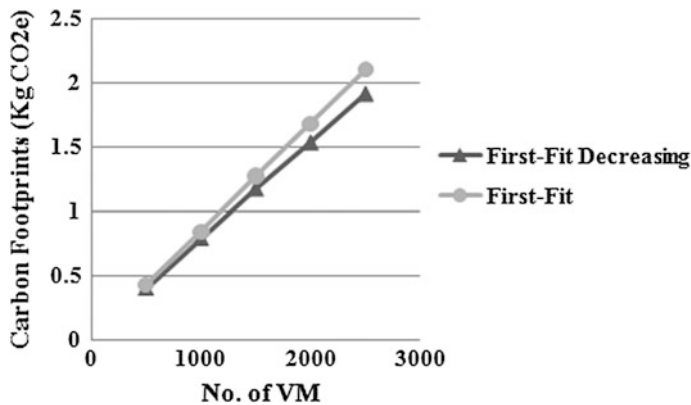


Fig. 6 Carbon footprints: first-fit decreasing and first-fit

On comparing the energy consumption of first-fit decreasing and first-fit algorithm using minimum migration time policy, it is evaluated that first-fit decreasing algorithm consumes less energy than first-fit algorithm, as depicted in Fig. 6.

From Table 1, it is observed that by using minimum migration time policy with first-fit decreasing algorithm 84 % carbon footprints can be saved with respect to non environment-aware policy and giving 2.8 % SLA violation. While minimum migration time policy with first-fit algorithm saves 83 % carbon footprints with respect to non environment-aware policy and giving 0.32 % SLA violation.

6 Conclusion

Carbon footprints are estimated for the proposed architecture obtained by extending the existing classical cloud architecture. This novel architecture is obtained by adding an energy-aware service provider layer. Live migration applied in this proposed architecture minimizes the number of running physical machines to minimize the required energy. The standard bin-packing algorithms; first-fit decreasing and first-fit used along with live migration save 84 and 83 % carbon footprints, respectively. As a future work, impact on other factors like propagation delay, server's temperature, etc., can be studied. Virtual machine allocation can be done using other resources like RAM, disk, bandwidth, etc., instead of using only CPU utilization.

References

1. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. Technical Report. National Institute of Standards and Technology, USA (2011).
2. Delforge, P., Whitney, J.: Data Center Efficiency Assessment. Technical Report. Natural Resource Defence Council, New York (2014).
3. Calheiros, R. N., Ranjan, R., Beloglazov, A., Rose, C. A. F. D., Buyya, R.: CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software- Practice & Experience*. vol. 21, issue. 1. John Wiley & Sons Ltd, New York (2011).
4. Department for Environment Food & Rural Affairs, <http://www.ukconversionfactorscarbonsmart.co.uk/>.
5. Verma, A., Ahuja, P., Neogi, A.: pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems. In: Issarny, V., Schantz, R. (eds.) *Middleware 2008*. LNCS, vol. 5346, pp. 234–264. Springer, Heidelberg (2008).
6. Buyya, R., Beloglazov, A.: Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers. *Concurrency and Computation: Practice and Experience*. vol. 24, issue. 13, pp: 1397–1420. John Wiley & Sons Ltd, UK (2012).
7. Rajabi, A., Ebrahimirad, V., Yazdani, N.: Decision Support-as-a-Service: An Energy-aware Decision Support Service in Cloud Computing. In: *5th Conference on Information and Knowledge Technology*, pp. 71–76. IEEE Press, Iran (2013).
8. Li, Y., Wang, Y., Yin, B., Guan, L.: An Energy Efficient Resource Management Method in Virtualized Cloud Environment. In: *14th Asia-Pacific Network operations and management symposium*, pp. 1–8. IEEE Press, South Korea (2012).
9. Wang, J., Huang, C., He, K., Wang, X., Chen, X., Qin, K.: An Energy-aware Resource Allocation Heuristics for VM Scheduling in Cloud. In: *IEEE International Conference on High Performance Computing and Communications & IEEE 10th International Conference on Embedded and Ubiquitous Computing*, pp. 587–594. IEEE Press, China (2013).
10. Clark, C., Fraser, K., Hand, S., Hansen, J.G., Jul, E., Limpach, C., Pratt, I., Warfield, A.: Live Migration of Virtual Machines. In: *2nd USENIX Symposium on Network Systems Design and Implementations*, pp. 273–286. USENIX Association, Berkeley (2005).
11. Brucker, P.: *Scheduling Algorithm*, 5th ed., Springer-Verlag Berlin Heidelberg (2007).
12. Coffman, E.G., Csirik, J., Woeginger, G.J.: Approximate Solutions to Bin Packing Problems. In: *Applied Optimization*, Pardalos, P.M., Resende, M.G.C (eds.), *Handbook of Applied Optimization*, Oxford University Press, pp. 607–615, New York (2002).

Examining Usability of Classes in Collaboration with SPL Feature Model

Geetika Vyas and Amita Sharma

Abstract Software product line engineering paradigm focuses on developing families of products keeping track of their common aspects and predicted variabilities. Feature models are often used for depicting the commonalities and variabilities existing in software product lines. Classes are used to program the features of the feature models and thus a significant relationship exists between the two. As software product line focuses on reuse, we have proposed a metric to measure the degree of usability of classes in context of features which are using them. Eclipse FeatureIDE is used to prove the proposed metrics. The aim of the research is to track usability of classes keeping in mind their planned reuse, efficient development and maintenance.

Keywords Software product line engineering · Features · Feature models · Degree of usability · Eclipse FeatureIDE

1 Introduction

Software product lines engineering develops and maintains families of products keeping track of their common aspects and predicted variabilities [1]. It focuses on reusability [2]. It is structured into two main processes: domain engineering (also called engineering for reuse) and application engineering (engineering with reuse) [3]. Features are structures that extend and modify the structure of a given program in order to meet the user requirement. Feature models introduced by Kang are used to represent the features available in a product line. They portray all the configurations a product line can possibly have [4]. The concept of feature is useful for description of commonalities and variabilities not only in the analysis and design but also

Geetika Vyas (✉) · Amita Sharma
The IIS University, Jaipur, India
e-mail: geetika_vyas@yahoo.co.in

Amita Sharma
e-mail: amitasharma214@gmail.com

implementation phase of software product lines [5]. There exists a significant relationship between classes and features, but no significant work is done in reference to the complexity that exists across the feature–class relationship. A feature in a feature model is supported by class(s) in a class diagram. Software product line paradigm aims reuse; and like features, classes are also reused. Therefore, the relationship between feature model and class diagrams needs to be studied. The core focus of this paper is to investigate the usability of classes. We have proposed a metric to measure degree of usability of classes. Collaboration diagrams are used to check the result generated by the metrics. The proposed metric is beneficial from the point of view wherein we are able to detect the origin point of most vital classes in the feature model, and also detect the most vital features and the least vital features possibly turning dead in the future. Other possible benefits seen behind the research are planned usage of classes in the system, their better development followed by improved maintenance. The rest of the paper is organized as follows: Section 2 contains introduction of feature-oriented programming. Section 3 introduces Eclipse FeatureIDE. Section 4 contains the proposed metrics and its implementation. Section 5 contains the result, analysis and conclusion.

2 Feature-Oriented Programming

Feature-oriented programming paradigm allows decomposition of a program into its constituent features. It was designed for software product line paradigm that allows significant code reuse and the generation of many similar but functionally different programs from the same set of features simply through selection of desired features [6]. The stepwise refinement leads to a layered stack of features. This helps in constructing well-structured software that can be tailored to the specific needs of the user and the application scenario [7, 8].

3 FeatureIDE: Eclipse Plug-in

FeatureIDE is an eclipse-based integrated development environment (IDE). It provides tool support for the feature-oriented design process and implementation of software product lines [9]. Eclipse FeatureIDE provides the most powerful and commercially successful open-source enhanced IDE support for feature-oriented programming implementations [10]. Domain analysis and feature modeling are supported with graphical feature model editor. Feature implementation is supported by variety of composers like AHEAD, FeatureC++, FeatureHouse, AspectJ, DeltaJ, Munge and Antenna building program families. Out of these we have used FeatureHouse which is language independent.

4 Experimental Setup

4.1 Implementation of the Proposed Metrics

In our Previous paper we have proposed metric for degree of usability [11] let us assume an anonymous feature model and implement the proposed metrics over it. Figure 1 contains this feature model, where F1 is the root node. It has three children: F2 (mandatory), F3 (mandatory) and F4 (optional). The parent node F4 has two child features: F5 and F6. Parent nodes F2 and F3 have one mandatory child each F7 and F8, respectively. There are following dummy classes, C1, C2, C3, C4 and C5, used to implement this feature model. The usage of these classes by the features is shown in Table 1.

The degree of usability can be defined as the number of times a class is used in different features present in a feature model across the tree. It is obvious that at the root node degree of all the classes will be zero, i.e. at the origin of the class its degree of usability will always be zero. Irrespective of the traversal method the final value of degree of usability of any class will always remain same. Table 1 displays the individual class usage scenario across the feature model. It also shows the calculated value of degree of usability following both methods of traversal, i.e. breath first and depth first.

On the basis of the calculations in the above table we can conclude that classes 1 and 3 have the highest usability. They are used maximum number of times, in comparison to the other classes. The value obtained by this metric is of great worth because it is an indicator of their usage highlighting their importance and subsequent use. Degree of usability can also be derived by classifying abstract and concrete classes. The collaboration diagram generated for this feature model also reflects the same value of usability of each class across each feature. Figure 2 proves our metric, wherein it can be clearly seen that classes 1 and 3 have the maximum reusability.

4.2 Implementation of the Proposed Metric

For implementing the proposed metric, we take the example of the Direct-to-home (DTH) systems. To implement our metric we are taking the broader aspect of DTH.

Fig. 1 Anonymous feature model developed using eclipse

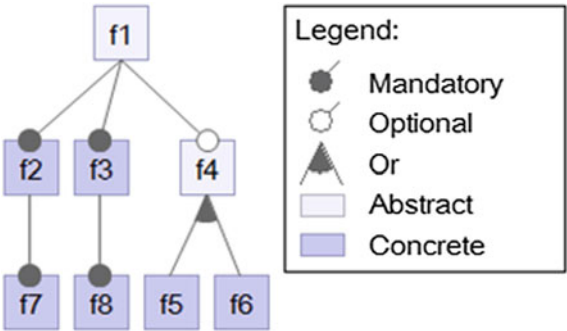
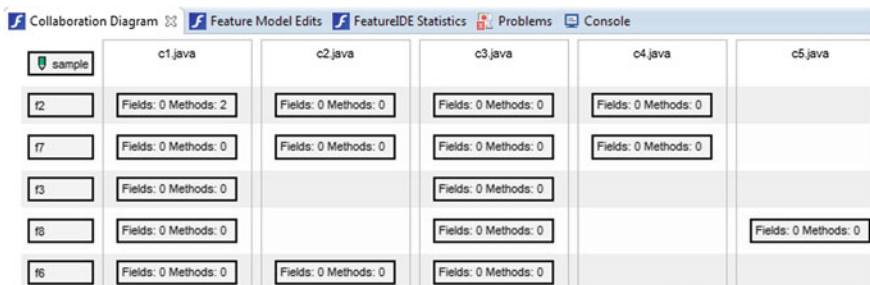


Table 1 Calculated values: degree of usability

Name of feature	Name of class	Degree of usability (breath first traversal)	Degree of usability (depth first traversal)
F1 (root feature)	–	–	–
F2	Class 1, 2, 3, 4	$d(C1) = 1$	$d(C1) = 1$
		$d(C2) = 1$	$d(C2) = 1$
		$d(C3) = 1$	$d(C3) = 1$
		$d(C4) = 1$	$d(C4) = 1$
F3	Class 1, 3	$d(C1) = 2$	$d(C1) = 4$
		$d(C3) = 2$	$d(C3) = 4$
F4	Class 1, 2, 4	$d(C1) = 3$	$d(C1) = 6$
		$d(C2) = 2$	$d(C2) = 5$
		$d(C4) = 2$	$d(C4) = 4$
F7	Class 1, 2, 3, 4	$d(C1) = 4$	$d(C1) = 2$
		$d(C2) = 3$	$d(C2) = 2$
		$d(C3) = 3$	$d(C3) = 2$
		$d(C4) = 3$	$d(C4) = 2$
F8	Class 1, 3, 5	$d(C1) = 5$	$d(C1) = 3$
		$d(C3) = 4$	$d(C3) = 3$
		$d(C5) = 1$	$d(C5) = 1$
F5	Class 2, 3, 4	$d(C2) = 4$	$d(C2) = 2$
		$d(C3) = 5$	$d(C3) = 5$
		$d(C4) = 4$	$d(C4) = 3$
F6	Class 1, 2, 3	$d(C1) = 6$	$d(C1) = 5$
		$d(C2) = 5$	$d(C2) = 4$
		$d(C3) = 6$	$d(C3) = 6$

**Fig. 2** Collaboration diagram for anonymous feature mode

We are focusing on its limited functionality and services. This television service is the reception of satellite programs with a personal dish installed individually at home. Its network consists of modulators, broadcasting center, encoders, satellites, multiplexers and DTH receivers. Here service provider leases Ku-band

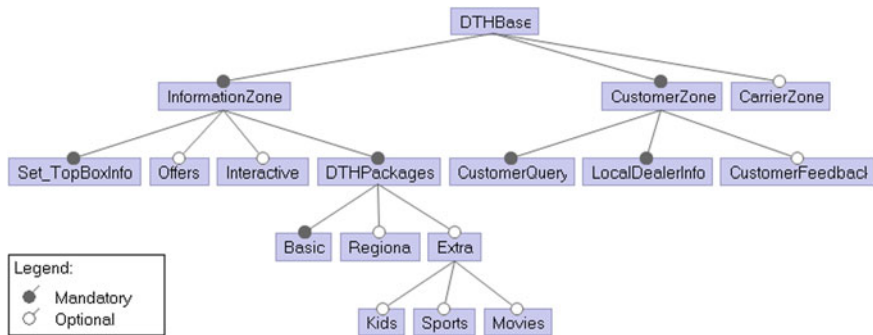


Fig. 3 Contains the feature model

transponders from the satellite. The audio, video and data signals are converted into the digital format and the multiplexer mixes these signals. At the users end, there is a small dish antenna installed and set-top boxes to decode it and viewing of numerous channels. The smallest receiving dish can be 45 cm in diameter. This transmission travels directly to the consumer through a satellite. DTH also offers stereophonic sound effects. Its advantage is that it can also reach remote areas where terrestrial transmission and cable TV cannot penetrate. Along with enhanced picture quality, other benefits are that it allows interactive TV services such as movie-on-demand, internet access, video conferencing and e-mail also. Figure 3 shows the DTH feature model.

Here DTHBase (root feature) has InformationZone (mandatory), CustomerZone (mandatory) and CarrierZone (optional) features. InformationZone has two mandatory features and two optional features, out of which DTHPackages feature has Basic (mandatory) and two features Regional (optional) and Extra (optional) features. Feature Extra has Kids (optional), Sports (optional) and Movies (optional) features. In total this feature model can have 144 valid configurations.

The basic (dummy) classes in this software include CostInfo, CustInfo, LocalDealerInfo and PackageInfo. The java files which use these dummy classes are jak files (extended files of java), also called FeatureIDE files. In later stages, as per need these classes will be refined in order to add new features in the software. These classes are dummy by nature. Implementation needs more effort on the programmer's part.

Using the depth first traversal method, the degree of usability of dummy class CostInfo is as follows:

At Feature InformationZone, $d(\text{CostInfo}) = 1$, (assuming the degree of Class CostInfo1 at DTHBase is 0)

At Feature Set_TopBoxInfo, $d(\text{CostInfo}) = 2$,

At Feature DTHPackages, $d(\text{CostInfo}) = 3$,

At Feature Basic, $d(\text{CostInfo}) = 4$,

At Feature CustomerQuery, $d(\text{CostInfo}) = 5$.

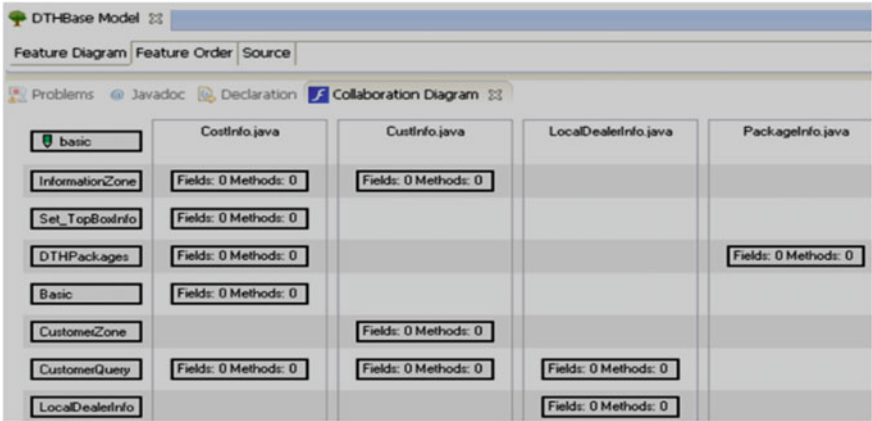


Fig. 4 Collaboration diagram of DTH service feature model

Using the breath first traversal method, the degree of usability of dummy class CostInfo is as follows:

- At Feature InformationZone, $d(\text{CostInfo}) = 1$, (the degree of Class CostInfo1 at feature DTHBase is 0)
- At Feature Set_TopBoxInfo, $d(\text{CostInfo}) = 2$,
- At Feature DTHPackages, $d(\text{CostInfo}) = 3$,
- At Feature CustomerQuery, $d(\text{CostInfo}) = 4$,
- At Feature Basic, $d(\text{CostInfo}) = 5$.

Thus we can conclude that the degree of usability of dummy class CostInfo, irrespective of the traversal method, is 5 and is the highest. To check whether the metric is returning the correct value, we refer to the collaboration diagram generated by Eclipse FeatureIDE. Once we define the FeatureIDE files, FeatureIDE generates a collaboration diagram which shows the collaboration of all classes with feature. Figure 4 contains the collaboration diagram for this example. The columns in the diagram contain the classes and rows contain the features which are using these classes. It clearly depicts that class CostInfo is the most referred class. Out of the four dummy classes it is the most frequently used one. Through the diagram also we come to the conclusion that the degree of usability of dummy class CostInfo is 5.

5 Analysis and Conclusion

A significant relationship is seen between features and classes. The strong association between these two leads us to relate the core focus of SPL in both the respects, i.e. to discuss usability of features and classes as well. The available measures in literature limit the complexity within the features. The complexity across the classes and features relationship remains untouched. Available metrics do not suffice in

controlling the usability of the whole system. The metric proposed in our paper is generating the degree of usability of various classes used in the example of DTH services. The collaboration diagram of the example also proves that the metrics are returning values which are true from the practical point of view. The classes which have highest usability theoretically have the same usability practically also. The calculated value thus obtained by our metric will help us check the usage of each class. This will ultimately benefit the programmers, practitioners and researchers in better understanding of classes. It will also help in improved control and development of the product line. It will help determine the best ways for the maintenance of classes which are an integral part of the whole process. Our current work is generalized by nature and is in its initial stages. Our proposal still needs validation. We are currently working upon the theoretical and empirical validations by studying variety of feature models and the classes used for implementing them [12]. We will also apply the metric over more examples to calculate accurate results. Further experimentation will validate our work and help us draw the final conclusions.

References

1. Clements, P., Northrop, L., *Software Product Lines: Practices and Patterns*, Addison-Wesley, Boston (2001).
2. Montagud, S., Insfran, E., Abrahão, S., A systematic review of quality attributes and measures for software product lines, *Software Quality Control*, Vol. 20, Issue 3–4, pp. 425–486, (2012).
3. Pohl, K., Bockle, G., Linden, F., *Software Product Line Engineering: Foundations, Principles and Techniques*, Springer-Verlag New York, Inc. Secaucus, NJ, USA, (2005).
4. Kang, Kyo, C., Lee, Jaejoon, Kim, Kijoo, Kim, Jounghyun, G., EuiSeob, S., Moonhang, H., FORM: A Feature-Oriented Reuse Method with Domain-Specific Reference Architectures. *Annals of Software Engineering* (Springer Netherlands), vol. 5, pp. 143–168, (2004).
5. Bagheri, E., Gasevic, D., Assessing the Maintainability of Software Product Line Feature Models Using Structural Metrics, *Software Quality Journal*, vol. 19 (3), pp. 579–612, (2011).
6. Apel, S., Kastner, C., An Overview of Feature-Oriented Software Development, *J. Object Technology* (JOT), vol. 8, No. 5, pp. 49–84, (2009).
7. Sharma, A., Sarangdevot, S.S., Investigating the Application of Feature-Oriented Programming in the Development of Banking Software Using Eclipse-FeatureIDE Environment, *International Journal of Computer Science & Technology*, Vol. 2, Issue 1, pp 53–57, (2011).
8. Rumbaugh, J., Blaha, M., R., *Object Oriented Modeling and Design*, Prentice Hall, (1991).
9. Leich, T., Apel, S., Marnitz, L., Tool Support for Feature-Oriented Software Development - FeatureIDE: An Eclipse-Based Approach, *OOPSLA Workshop on eclipse technology eXchange (ETX)*, San Diego, USA, (2005).
10. Kastner, C. FeatureIDE: A Tool Framework for Feature-Oriented Software Development, *Proc. 31st Int'l Conf. on Software Engineering (ICSE)*, Vancouver, Canada, IEEE, (2009).
11. Sharma, A., Vyas, G., New Set of Metrics for Accessing Usability in Feature Oriented Programming, *International Journal of Computer Applications*, vol. 81(11), pp. 19–22, (2013).
12. Siegmund, J.; Siegmund, N.; Apel, S., “Views on Internal and External Validity in Empirical Software Engineering,” in *Software Engineering (ICSE)*, 2015 IEEE/ACM 37th IEEE International Conference on, vol. 1, no., pp. 9–19, 16–24 May 2015.

Do Bad Smells Follow Some Pattern?

Anubhuti Garg, Mugdha Gupta, Garvit Bansal, Bharavi Mishra
and Vikas Bajpai

Abstract Software maintenance is a daunting task but equally crucial for an aging software. Software maintainability is one of the important quality aspects, which is directly affected by code smells. Software maintenance requires considerable amount of budget which is sometimes even much higher than the actual cost of software development. Some bad practices, such as code clones, anti-patterns, and bad smells, ultimately result in severe maintenance consequences. In this paper, an experimental attempt is made, which is based on market basket analysis to answer this question: “Whether bad smells follow some pattern or not?” by studying the behavior of bad smells and their co-occurrences.

Keywords Bad smell · Software maintenance · Market basket analysis

1 Introduction

Developing strategies for assessing the maintainability [1] of a system is of vital importance. Good design quality of software eases the non-functional attributes [2] such as maintainability, re-usability, flexibility, understandability, functionality, and extendibility. To achieve this in a better fashion one needs to understand various

Anubhuti Garg (✉) · Mugdha Gupta · Garvit Bansal · Bharavi Mishra · Vikas Bajpai
Department of Computer Science and Engineering, The LNM Institute
of Information Technology, Jaipur, India
e-mail: anubhuti.grg@gmail.com

Mugdha Gupta
e-mail: mugdhagupta4@gmail.com

Garvit Bansal
e-mail: garvitbansal244@gmail.com

Bharavi Mishra
e-mail: bharavi@lnmiit.ac.in

Vikas Bajpai
e-mail: vikas.bajpai87@gmail.com

code smells and their capacity to effect maintainability of the software. Code smells [3] are structural characteristics of software that can make software hard to evolve and maintain. The presence of code smells may degrade quality attributes, such as understandability and changeability which have significant effect on the performance of software. During experimental analysis, it is observed that several code smells tend to occur together and have some co-occurrence patterns. The interaction effects between various code smells can intensify problems caused by individual code smells or can lead to additional, unforeseen maintenance issues.

There has been considerable work done on the effect of LOC on bad smell (we have used bad smell and code smell interchangeably in the paper). Zhang et al. [4] investigate the functional form of the size–defect relationship for software modules. This facilitates various development decisions related to prioritization of quality assurance activities. In another work, Hui et al. [5] explored the detection and resolution sequences of different kinds of bad smells. Although these studies are useful, conducting them for large-scale study has always been a challenge. The primary challenge is the data collection as there is no tool available that can find all bad smells together. Saini et al. [6] have found bug patterns in component repositories. In another work of its kind, Yamashita et al. [7–9] investigated the comprehensive and informative nature of code smells that can be deployed for the assessment of software maintainability.

Software companies like Mozilla Firefox, Chromium, Google, and many others keep releasing new versions because of such defects in the code. In this paper, empirical study is performed to investigate the hierarchical relationship of bad smells on two open-source softwares namely Mozilla and Chromium. We attempt to answer the following research question:

- (1) Is there any pattern (positive or negative) exists in the occurrences of various code smells?

The rest of the paper is organized as follows: it starts by giving details of the bad smells that were found and the tools used for finding them in Sect. 2. Then Sect. 3 describes the research methodology and data collection. Section 4 discusses the results and relevant information that can be extracted by showing the hierarchical relationship of bad smells in package and classes. Finally, concluding remarks are made in Sect. 5.

2 Code Smells: Preliminaries

Bad smells are signs of potential problems in code. They do not currently prevent the working of the program from functioning; however, they indicate weakness in design and coding of the software. It is essential to retain the software quality because of our increasing dependency. Beck and Fowler et al. [10] provided a set of informal descriptions of 22 code smells and associated them with different re-factoring

strategies that can be applied to improve software design. Here are the definitions of some code smells on which we have focused our attention in this paper.

- *Data Class*—A class whose purpose is to hold data has instance variables, getters, and setters methods.
- *God Class*—A class takes too many responsibilities. It centralizes the system functionality in one class, which contradicts the decomposition design principles.
- *God Module*—God module is an abnormally large, complex, and non-cohesive module (i.e., a file containing global functions and variables), which excessively manipulates global variables exposed by other modules.
- *Data Module*—Data module describes a module (i.e., a file containing global functions and variables) that exposes too much of its global variables instead of providing global functions. It is the procedural equivalent of a data class.
- *Feature Envy*—Feature envy refers to an operation that is manipulating a lot of data external to its definition scope. In object-oriented code this is a method that uses many data members from a few other classes, instead of using the data members of its own class.
- *Data Clumps*—Data clumps are large groups of parameters that appear together in the signature of many operations.
- *Code Duplication (Internal/External)*—Code duplication refers to groups of operations which contain identical or slightly adapted code fragments. By breaking the essential don't repeat yourself (DRY) rule, duplicated code multiplies the maintenance effort, including the management of changes and bug fixes. Based on the different re-factoring approaches, it is distinguished into *internal duplication* (involving methods that belong to the same scope, i.e., class or module) and *external duplication* (that refer to unrelated operations).

We have used software *InCode* to detect bad smells at package level. *InCode Helium* is an open-source software developed by *Intooitus*. It is used as a quality assessment tool for the codes written in Java, C++, and C. It detects design flaws automatically and helps to understand the causes of quality problems on the level of code and design.

3 Research Methodology

In this study, market basket analysis [11, 12] is used to study the behavior of bad smells and their co-occurrences. Market basket analysis is a modeling technique which predicts the future behavior of person, product, or software using their associative information. *Bitmap* approach, an older version of association mining technique, is applied to investigate the *positive as well as negative co-occurrences* of bad smells in the software. The term *positive co-occurrence* refers to the existence of particular combination of bad smell, and *negative co-occurrence* refers to the non-existence of particular combination of bad smell. Traditional Apriori

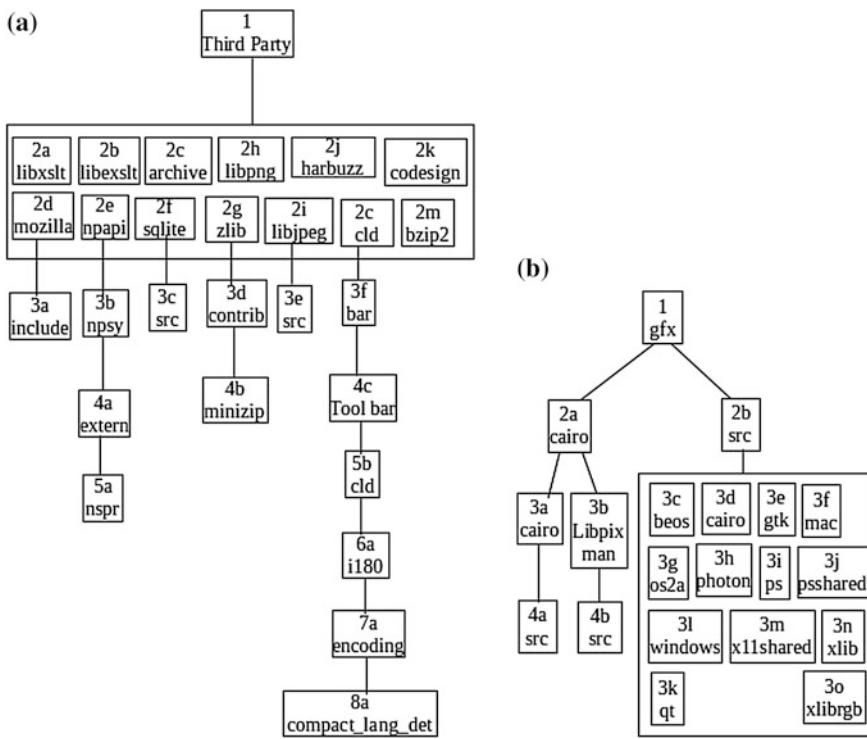


Fig. 1 a Chromium. b Mozilla

association rule mining algorithm requires having domain knowledge toward setting the support and confidence parameters. In bitmap approach, a pool draining fundamental of rule fishing is used. For each of the observing softwares, a bitmap is generated for corresponding high level of packages using InCode tool. In bitmap, a cell $C[P, f_i]$ contains '1' if package P contains bad smell f_i ; otherwise contains '0'. Different levels of granularity of co-occurrences or number of item sets are used to analyze the associative relationship of bad smells. Two different types of analysis are performed on two softwares: (1) Software as a whole and (2) At package level. We used one package for each software with considerable high LOC for package-level analysis of *Third_party* for Chromium and *Gfx* for Mozilla. Hierarchical package-level relationships of the software are depicted in Fig. 1a for *Third_party* of Chromium and Fig. 1b for *Gfx* for Mozilla. Each of the levels in the graph depicts the package-level abstraction of the software. Each node at corresponding level has two labels: label index provided as 1; 2a, 2b...; 3a, 3b...; 4a, 4b...; so on and package name. A directed edge between two nodes $v1$ and $v2$ represents hierarchical derivation of package $v2$ from higher level package $v1$. There are eight levels of abstraction available for *Third_party* package and three

levels of abstraction for *Gfx* package. To avoid the complexities we only considered the leaf nodes at each level and discarded all the internal nodes while analyzing the co-occurrences of bad smells.

4 Results

Bad smell detection is a worthy process in software maintenance prospective. By analyzing Table 1, it is observed that both the softwares do not contain any contradictory patterns. Most of the bad smells patterns are associated with both the modules with some magnitude of differences.

Table 1 Co-occurrence of bad smells at abstract level

Code smell	Code smell	Co-occurrence (%) in Chromium software	Co-occurrence (%) in Mozilla software
Data class	God class	43.7	58.9
Data class	God module	3.12	17.9
Data class	Data module	21.8	28.2
Data class	Feature envy	46.8	71.79
Data class	Data clumps	43.7	71.7
Data class	Internal duplication	28.1	69
Data class	External duplication	15.6	61.5
God class	God module	3.12	15.3
God class	Data module	15.6	23
God class	Feature envy	40.6	58.9
God class	Data clumps	31.2	51.2
God class	Internal duplication	21.8	56.4
God class	External duplication	12.5	48
God module	Data module	3.12	15
God module	Feature envy	3.12	17.9
God module	Data clumps	3.12	17.9
God module	Internal duplication	3.12	17.9
God module	External duplication	3.12	17.9
Data module	Feature envy	18.75	25.6
Data module	Data clumps	18.75	28.2
Data module	Internal duplication	18.75	25.6
Data module	External duplication	15.6	28.2
Feature envy	Data clumps	34.33	61.5
Feature envy	Internal duplication	25	64.1
Feature envy	External duplication	15.6	53.8
Data clumps	Internal duplication	21.8	66.6
Data clumps	External duplication	15.6	58.9
Internal duplication	External duplication	18.75	61.5

Table 2 Co-occurrence of bad smells in Chromium

	Code smell	Occurrence in leaf nodes
Two code smells	(Data clumps, internal duplication)	2a, 2m, 3e, 8a
	(Internal duplication, external duplication)	2a, 2h, 2k, 3e
	(Data clumps, external duplication)	2a, 3c, 3e
	(Feature envy, internal duplication)	2h
	(Feature envy, external duplication)	2h
Three code smells	(Data clumps, internal duplication, external duplication)	2a, 3e
	(Feature envy, internal duplication, external duplication)	2h

Some patterns appear with high percentage of occurrence in both the softwares. For instance, (Data Class, God class), (Data Class, Data Module), etc. are occurred with high percentages in both the softwares. In contrast, some patterns such as (God Class, God Module), (Data Class, God Module), etc. occurred with high level of percentage differences. We did not find any negative pattern of bad smells in both the softwares.

Package-wise co-occurrences of bad smells are shown in Tables 2 and 3 for Chromium and Mozilla, respectively. The analysis of Tables 2 and 3 indicates that most of the leaf level packages contain bad smells and bad smell patterns. To reveal the contribution of bad smells we also analyzed the results in different levels of abstraction. In Mozilla, at lowest level (4a and 4b) only three classes of bad smells appear as data clumps, internal duplication, and external duplication. These two packages contain 3a and 3b and at this level again the same classes of bad smells are detected with high level of magnitude. At the second level and top levels ((2a, Cario) and (1, *Gfx*)) same pattern is repeated which therefore indicates that bad smells are repeating themselves in subsequent abstraction level (Tables 4 and 5).

Table 3 Co-occurrence of bad smells in Mozilla

	Code smell	Occurrence in leaf nodes
Two code smells	(Data clumps, internal duplication)	3e, 3f, 3n, 3o, 4a, 4b
	(God class, internal duplication)	3c, 3e, 3h, 3i, 3n
	(God class, feature envy)	3d, 3e, 3i, 3l
	(God class, data clumps)	3e, 3g, 3m, 3n
	(Data clumps, external duplication)	3e, 4a, 4b
	(Feature envy, internal duplication)	3e, 3i
	(Internal duplication, external duplication)	4a, 4b
	(God class, external duplication)	3e
Three code smells	(God class, feature envy, internal duplication)	3e, 3i
	(Data clumps, internal, external duplication)	4a, 4b
	(God class, data clumps, internal duplication)	3n

Table 4 Level-wise distribution of bad smells in Mozilla

Package	2	3	4	File	Data clump	Internal duplication	External duplication
<i>Gfx</i>	2a	3a	4a	Cario-xlib-surface.c	10	3	2
				Cario-surface.c	9	2	
				Xcb-surface.c	5		2
				Path-data.c		2	
	2a	3b	4b	Fbcompose.c	66	1	
				Fbmmx.c	15	11	2
				Fbpict.c	16		2
				Icrti.c	3	3	

Table 5 Level-wise distribution of bad smells in Chromium

Package	2	3	4	5	6	7	8	Data class	Data clumps	Internal duplication
Third/party	2i	3f	4c	5b	6a	7a	8a	3	5	4
Third/party	2i	3f	4c	5b	6a	7a		3	5	4
Third/party	2i	3f	4c	5b	6a			4	5	4
Third/party	2i	3f	4c	5b				4	5	4
Third/party	2I	3f	4c					5	5	4
Third/party	2i	3f						5	5	4
Third/party	2i							8	5	4
Third/party								801	284	866

In Chromium software, there are eight levels of abstraction. At the lowest level only data clumps and internal duplication are detected. At the seventh level one more bad smell data class is detected. From level seven to level one, no new bad smell is detected; only their numbers are increased slightly. After analyzing Third_party as a whole package, it is observed that the number of bad smells at the lower level significantly contributes to the upper level. It is also observed that data clumps, internal duplication, and external duplication are the base bad smells and contributed significantly in making maintenance most crucial task in software development.

5 Conclusions

The goal of this study is to investigate the co-occurrence of bad smells in open-source software. We used bit string approach on two open-source softwares, Mozilla and Chromium, to detect the bad smell patterns and hierarchal relationship of their

occurrence at different levels of abstractions. It has been observed that co-occurrence patterns are presented in both the softwares with slight variation in their co-occurrence percentage. Some bad smells are more common such as *data clumps*, *internal duplication*, and *external duplication* and contributed significantly in subsequent abstract level. We can reduce the impact of bad smells or contribution of bad smells on subsequent upper level by analyzing and applying the detection algorithm at the file level because undetected bad smells at the lower level are combined to produce more serious and sophisticated bad smells. In this research we have worked on a single version of two softwares. We henceforth plan to extend our work to detect the persistent bad smells and their impact on quality assurance activity.

References

1. G. Parikh and N. Zvegintzov, "Tutorial on Software Maintenance", IEEE CS Press, Los Alamitos, Calif., Order No. 453, 1983.
2. Glinz, Martin. "On non-functional requirements." Requirements Engineering Conference, 2007. RE'07. 15th IEEE International. IEEE, 2007.
3. M. Mantyla, "Bad smells in software - a taxonomy and an empirical study." Ph.D. dissertation, Helsinki University of Technology, 2003.
4. Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining Association rules between sets of items in large databases" Proceedings of the 1993 ACM SIGMOD International conference on Management of data, SIGMOD '93, pg-207.
5. Vaibhav Saini, Hitesh Sajnani, Joel Ossher, Cristina V. Lopes et al. "A Dataset for Maven Artifacts and Bug Patterns Found in Them", MSR '14, May 31-June 1, 2014.
6. Hui Liu, Zhiyi Ma, Weizhong Shao, and Zhendong Niu et al "Schedule of Bad Smell Detection and Resolution: A New Way to Save Effort". In IEEE Transaction on Software Engineering, Vol. 38, No. 1, January/February 2012.
7. Aiko Yamashita, "Assessing the Capability of Code Smell to Support Software Maintainability Assessments: Empirical Inquiry and Methodological Approach." In Doctoral thesis, June 2012.
8. Aiko Yamashita, Leon Moonen et al. "Exploring the impact of inter-smell relations in the maintainability of a system: An empirical study." In ICSE'13 Proceedings of the 2013 International Conference on Software Engineering, pg 682–691.
9. Aiko Yamashita, Leon Moonen et al. "Do code smells reflect important maintainability aspects?" In 28th IEEE International Conference on Software Maintenance (ICSM), 2012.
10. Kent Beck, Martin Fowler et al. "Refactoring: Improving the Design of Existing Code." Addison-Wesley Longman Publishing Co. Inc., Boston, MA, USA, 1999.
11. A. Gunes Koru, Dongsong Zhang, Khaled El Emam and Hongfang Liu et al. "An Investigation into Functional Form of Size-Defect Relationship for Software Modules." In IEEE Transaction on Software Engineering, vol. 35, no. 2, March/April 2009.
12. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen & A.I. Verkamo, 1996. "Fast discovery of association rules. Advances in Knowledge Discovery and Data Mining" (U. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy, ed.), American Association for Artificial Intelligence.

Ultrawideband Antenna with Triple Band-Notched Characteristics

Monika Kunwal, Gaurav Bharadwaj, Kiran Aseri and Sunita

Abstract Nowadays, world has been moving toward augmented data rate and performance of antenna. UWB has been adapted due to its higher data rate over the large bandwidth. A compact ultrawideband antenna with triband rejection characteristics has been proposed. The three types of notches can be obtained by inserting two slots in the ground structure and one slot in the radiating patch, respectively. The proposed antenna not only shows better radiation pattern but also provides constant gain over the ultrawideband with the exception of notched frequency band. CST Microwave Studio software is used for optimizing the parameters of UWB antenna with band-notch features.

Keywords Band reject antenna · UWB antenna

1 Introduction

Owing to the progress in the field of wireless communication, UWB antenna has procuring more attention because of augmented data rate, little power emission, compact in size, low profile, omnidirectional radiation pattern, inexpensive, little power consumption, high radiation efficiency, low group delay, high security, and low cost. In 2002, the Federal Communication Commission permitted the unlicensed band that starts from 3.1 to 10.6 GHz for ultrawideband application [1]. In this frequency range, various other systems share the same bandwidth and thus they

Monika Kunwal (✉) · Gaurav Bharadwaj · Kiran Aseri · Sunita
Government Women Engineering College, Ajmer, India
e-mail: monikakunwal@gmail.com

Gaurav Bharadwaj
e-mail: gwecaexam@gmail.com

Kiran Aseri
e-mail: kiranrids18@gmail.com

Sunita
e-mail: olasunita30@gmail.com

create interference, so regulation is made on the UWB systems. The main center of attraction of UWB antenna is that it is not only easily fabricated on the PCB but also incorporated in the portable devices.

Several UWB antennas have been reported in the literature such as triangular, square, rectangular, circular, annular, elliptical, and hexagonal, in shape [2, 3]. Several single and multiband-notched antennas have been reported in the literature [1–6]. Various methods are available to get band-notched antenna which is used to etch the slot not only on the patch but also on the ground or on the feed [3–6].

In this paper, three band-rejecting antennas have been presented for UWB applications. C-shaped type slot is embedded in ground plane for eliminating band from 5.09 to 6.02 GHz; and for eliminating band from 2.352 to 2.67 GHz and from 3.118 to 3.76 GHz, E-shaped type slot and inverted U-shaped type slot are introduced in the radiating patch. The required band-notched frequency can be realized by altering the horizontal and vertical lengths of the desired band-notched structure.

2 Antenna Structure Design

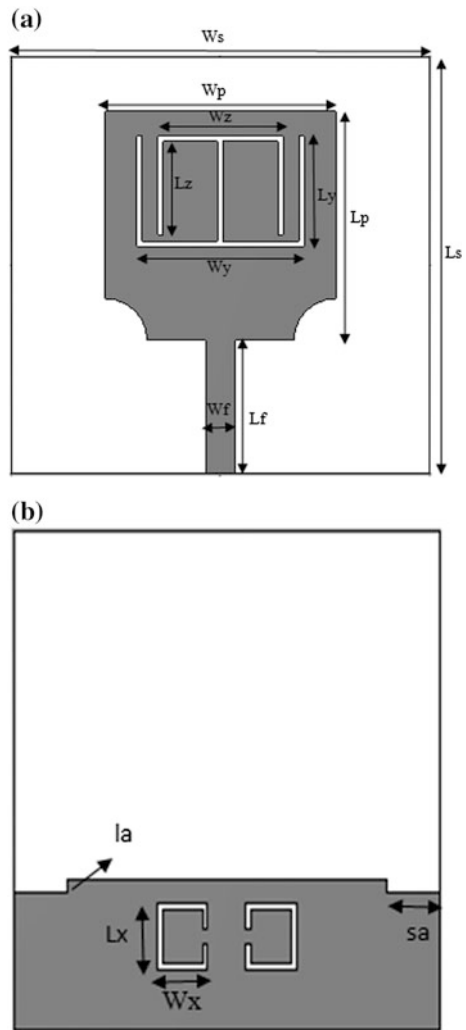
Figure 1 shows the configuration of UWB antenna with band-notch feature. The radiating patch is put on the FR-4 substrate with 1.6 mm thickness, 4.4 dielectric constant, and 0.02 loss tangent. 50 Ω microstrip line is used for feeding the antenna. The gap is introduced between the ground and the radiating patch for improving the VSWR. The dimensions of antenna structure are optimized using the software called CST in order to achieve better impedance bandwidth and to get stable gain and radiation characteristics (Table 1).

3 Result and Discussions

The simulated VSWR is shown in Fig. 2. The simulated impedance bandwidth of the UWB antenna is 2.25–10.3 GHz, for $S_{11} \leq -10$ dB. There are three stop bands in the frequency ranges from 2.363–2.792, 3.254–3.76, and 5.047–5.99 GHz, for VSWR > 2. Therefore, these stop bands are used to avoid interference with 2.5/3.5 GHz Wi-MAX and 5.5 GHz WLAN band.

For understanding the band-notched characteristics, the distribution of surface current or H field of UWB antenna at the center of the band-notched frequency has been investigated. The simulated surface current is mainly distributed around the E-shaped slot at the 2.51 GHz. At 3.51 and 5.509 GHz, surface current is mainly concentrated at edges of inverted U-shaped type slot and C-shaped type slot (Fig. 3).

Fig. 1 Configuration of the proposed antenna. **a** Top view. **b** Bottom view



The simulated gain and radiation efficiency of UWB antenna with triple band-notch features are shown in Figs. 4 and 5. Stable gain is obtained throughout the UWB band except at the rejection band. Almost 80 % radiation efficiency is obtained throughout UWB band except the rejected bands.

Figure 6 illustrates the *E*- and *H*-plane patterns of UWB antenna with triple band notch at 3, 5.5, and 10 GHz. At lower frequency (i.e., 3 GHz), the radiation pattern is like a dipole and at higher frequency (i.e., 10 GHz), the radiation pattern has many lobes in the *E*-plane.

Table 1 Dimension of the antenna structure

Parameters	Values (mm)
Ws	40
Ls	40
Wp	22
Lp	22
Wf	2.8
Lf	12.8
Wy	16
Ly	10.6
Wg	40
Lg	12
Wz	12
Lz	9.5
Wx	4.7
Lx	5.42
Hh	5
Sa	3
La	0.5

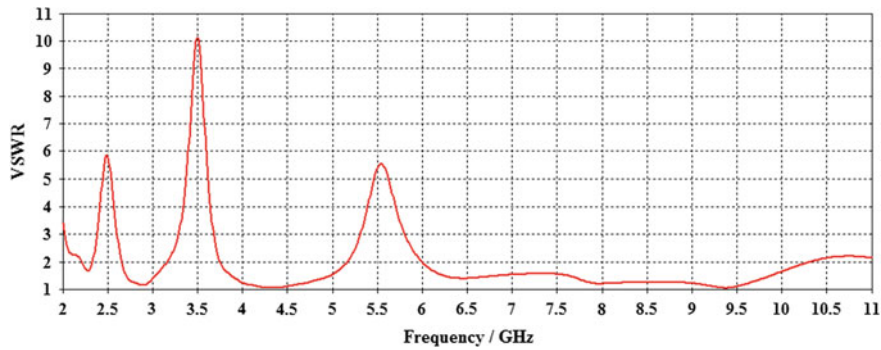


Fig. 2 The VSWR of the proposed antenna

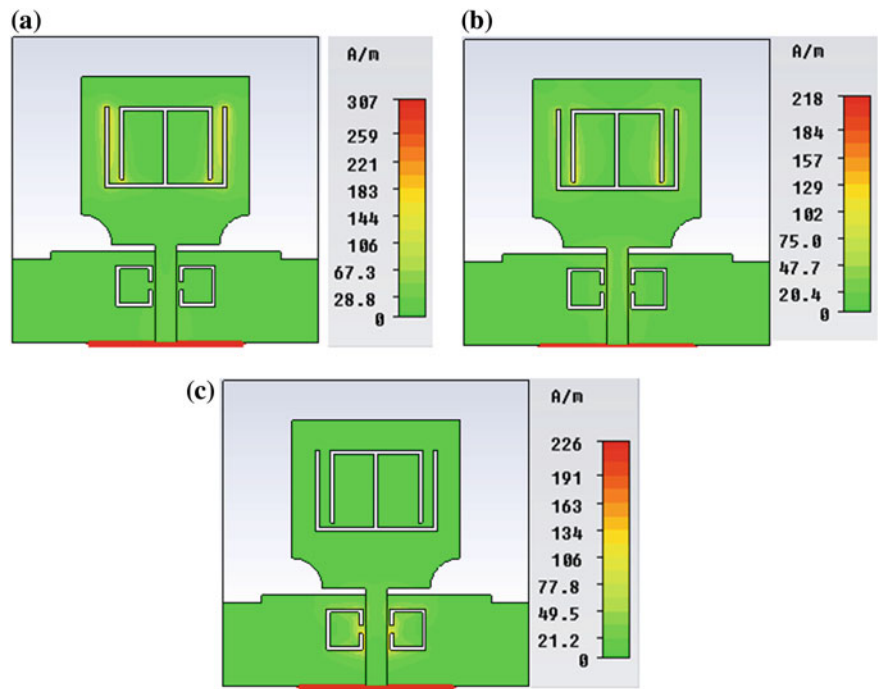


Fig. 3 The distribution of surface current of the proposed antenna at different frequencies

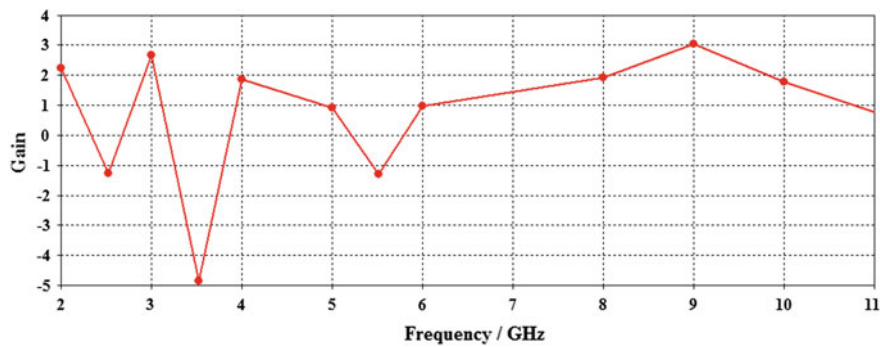


Fig. 4 The simulated gain of the proposed antenna

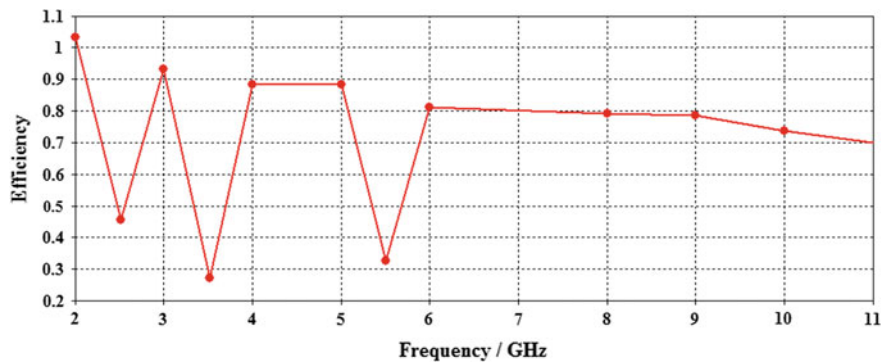


Fig. 5 The simulated radiation efficiency of the proposed antenna

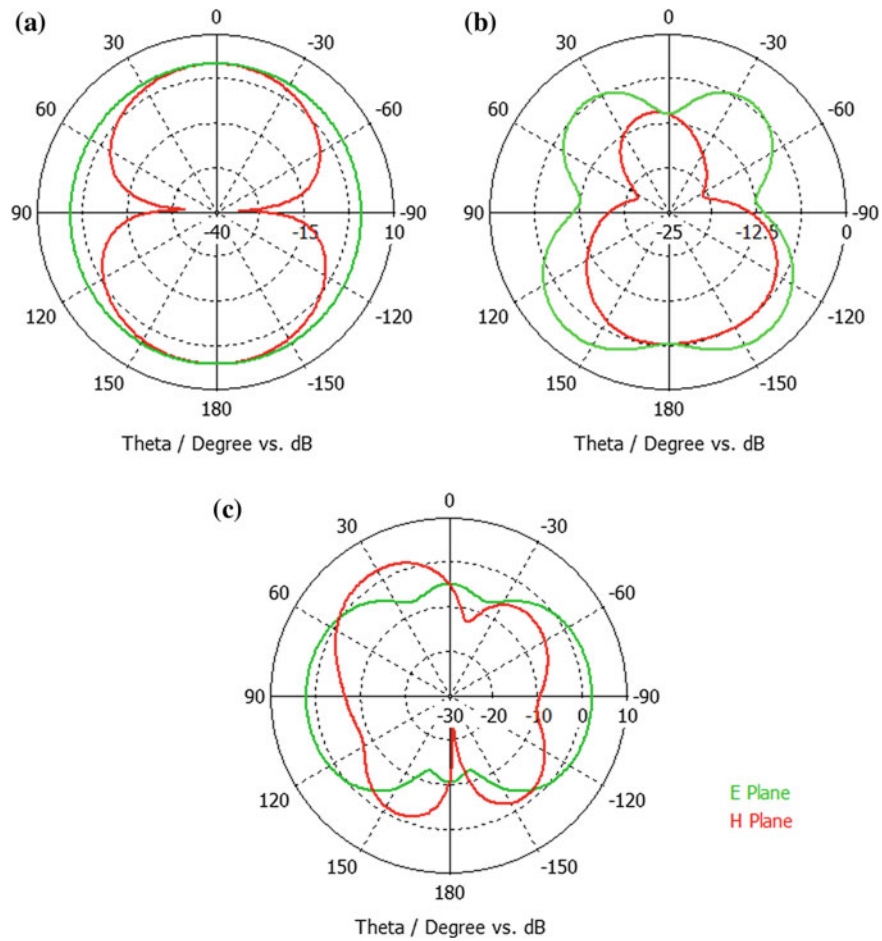


Fig. 6 The simulated radiation pattern of proposed antenna at different frequencies

4 Conclusion

Benefits of this antenna are easy to assemble, low cost, and simple structure. The fundamental frameworks of the antenna such as return loss, radiation patterns, and bandwidth are acquired. All frameworks satisfy the acceptable antenna standard and the satisfactory results are observed. The three stop bands are attained by introducing the E-shaped slot and the inverted U-shaped type slot and C-shaped type slot. The UWB antenna with triple band notch is expected to be good option to incorporate with UWB systems.

References

1. Tapan Mandal and Santanu Das: Ultrawideband-printed hexagonal monopole antennas with WLAN band rejection, *Microwave and Optical Technology Letters* Volume 54, Issue 6, pages 1520–1525 (2012).
2. Sanjeev K. Mishra. “Low-cost, compact printed circular monopole UWB antenna with 3.5/5.5-GHz dual band notched characteristics”, *Microwave and Optical Technology Letters* (2012).
3. J.M. Xi and C.H. Liang: CPW-fed trapezoidal antenna with dual band-notched characteristic for UWB application, *Microwave Opt Technol Lett* 52, pp. 898–900 (2010).
4. K. Yin and J.P. Xu: Compact ultra-wideband antenna with dual band stop characteristic, *Electron Lett* 44, pp. 453–454 (2008).
5. S. Natarajamani, Santanu Kumar Behera and Sarat Kumar Patra: A triple band-notched planar monopole antenna for ultrawide band applications, *Microwave and Optical Technology Letters* Volume 54, Issue 2, pages 539–543 (2012).
6. J.R. Panda, and R.S. Kshetrimayum: A compact CPW-fed monopole antenna with an U-shaped slot for 5 GHz/6 GHz band-notched ultrawideband applications, *Indian Antenna Week*, pp. 1–4 (2010).

Utilizing NL Text for Generating UML Diagrams

Prasanth Yalla and Nakul Sharma

Abstract UML diagrams form an important part of the software design specification. The source of these diagrams is requirement specification which is created from the user's need and requirements. In our work, we identify that two important areas in computer science and engineering, software engineering (SE) and natural language processing (NLP), form the core of this development. An algorithm for undertaking study of this approach is also presented. Herein, we also list the main usage of our technique to handle a more generalized environment such as non-software engineering domain.

Keywords Software engineering · Natural language processing · UML diagrams · Computational linguistics

1 Introduction

UML diagrams are currently in version 2.0 [1]. UML diagrams are within the realm of software development life cycle (SDLC) developed at the time of analysis and design phases of SDLC. They are formed when the analysis phase is about to get over and design phase is starting [1]. UML diagrams are generally developed manually although some attempts have been made to develop these diagrams from natural language text [2–5].

In the previous works undertaken, direct text has been utilized by scanning the relevant information in generating the UML diagrams [2, 3, 5]. However, if the input is better or suitable for the automation tools, then quality of the diagrams

Prasanth Yalla (✉) · Nakul Sharma
Computer Science and Engineering Department, K.L. University,
Guntur, India
e-mail: prasanthyalla@gmail.com

Nakul Sharma
e-mail: nakul777@gmail.com

which is generated can be improved. A textual use case description already exists in the form of meta-model [6] and use case template [7]. However, for other UML diagram such information is not available.

2 Problem Definition

The literature survey was undertaken keeping in view the software engineering and natural language processing tasks for this conversion.

Agt in his work has utilized the formal language sources to generate iterative approach for language engineers. The author develops an automated knowledge acquisition tool for supporting language engineering in the early phase of SDLC [8].

Stepane has developed a meta-model for textual use case description. The author utilizes the existing use case specification to generate a meta-model having OCL constraints [6].

Hause has discussed the role of use case diagrams outside the realm of software development. The author suggests role of use case in avionics system and system engineering. The pitfalls of use cases and the solutions are also presented [9].

Reed et al. described how different ontologies can be mapped onto Cyc. They have taken help of various subject experts in this work [10].

Simko et al. have made domain model from textual specification. The authors have utilized OpenNLP and CoreNLP technologies to complete this task [11].

Reynaldo et al. have developed class models through controlled requirements. The author accepts input as controlled N.L. text and validates with RAVEN project [2].

Sascha et al. have discussed how requirement engineering's error gets propagated to design and coding stages. The authors, hence, propose the automated analysis of N.L. text in SPIDER project [3].

More et al. have generated UML diagrams from N.L. text. The authors have utilized RAPID steaming algorithm and OpenNLP tools to accomplish this task [4].

Sudha et al. have described how natural language processing of tweeted text can help in times of crises. The authors have developed a classifier which can classify tweets for human analyses in times of crises [12].

Artis et al. have studied that UML models are inherently static. Hence, they have utilized a simulation environment called ARENA for running UML Models [13].

Mathias et al. have developed a requirement engineering feedback system (REFS) that checks for consistency with the textual requirement and models [14].

Bajwa et al. have developed class, activity, and sequence diagrams from simple english sentences. They have presented a methodology called UMLG to develop these UML diagrams. The authors claim that their algorithms can further be improved by introducing learning [15].

Bajwa et al. discussed an approach generating SVBR rules from natural language specification. The paper shows the importance automation in generating SVBR indicating that business analyst with loads of documents. They have developed an algorithm for detecting the semantics of English language [16].

Bajwa et al. highlighted the cases in which Stanford POS tagger does not identify the particular syntactic ambiguities in English specifications of software constraints. A novel approach to overcome these syntactic ambiguities is provided and better results are presented [17].

Bajwa et al. presented a new model for extracting necessary information from the natural language text. The authors generate use case, activity, class, and sequence diagram from the natural language text. The designed system also allows generation of system from natural language text [18].

Bajwa et al. proposed a SVBR approach to generate an unambiguous representation in English language. The input text is extracted for the relevant information of SVBR. A tool named NL2SVBRviaSBVR is made to accomplish this task [19].

Bajwa et al. proposed an interactive tool to draw use case diagrams. The authors have utilized LESSA approach for getting useful information from the natural language text [20].

The following research questions were not studied in-depth in the current literature, so some research questions were formulated as follows:

RQ-1. How is it possible to generate UML diagrams from natural language text?

RQ-2. Is it possible to develop a unified approach in developing UML diagrams from natural language text?

3 Problem Solution

To answer the following research questions, the following research strategy was adopted.

3.1 RQ-1

This question mainly dealt with the current state-of-the-art literature. In the literature review section, various papers dealing with this subject matter were studied. It is possible to generate UML diagrams in the following ways:

- Developing UML diagrams manually and then automating the developed diagrams.
This was the traditional way of developing UML diagrams. With this method, it is only possible to automate the resources which lead to developing the UML diagrams.
- Developing UML diagrams using NLP tools and techniques
This involves making use of NLP resources to scan the textual descriptions for getting information in generating UML diagrams [2, 3, 5, 6].

3.2 RQ-2

While trying to answer both these research questions, it is pertinent to understand how software engineering and natural language processing are interrelated to each other. Software engineering deals with how the software as a product will be engineered or made [1], while natural language processing (NLP) deals with utilizing machine or computer to better understand and process text or speech [12].

Type of UML diagram to be generated

There are the following types of specification while drawing UML diagrams:

- Unified specification,
- Booch specification,
- OMT specification.

These specifications only differ in the notation for various UML diagrams.

Generation of textual information

Textual information is very useful in understandability of any artifact in software development [21]. Textual description can aid in generating UML diagrams before any diagrams can be drawn [5].

Developing or using existing ontologies

In addition to plain text, ontologies in the form of heavy or light weight can help in generating good-quality UML diagrams [5, 11]. This helps in making the context clear especially when the plain text is not clear.

Human factors

The requirement engineer or designer who makes the UML diagrams must be well-versed with using the technical know-how of the software and the domain-level information. The project can be executed successfully when there is clarity in the UML diagrams developed by the humans [1]. Since the evaluator of any UML diagram is ultimately human, it is important that human factors such as understanding are also taken into account. This includes the know-how of person developing and using the system [1].

Issues at level of natural language processing

Level of noise in a sentence

The sentence which forms input to the text of UML diagram must be free from the noise [2]. The sentence must, hence, be scanned with appropriate tool which gives noise-free sentences.

Determining the complexity of sentence

An algorithm needs to be designed to check the complexity of a sentence. A benchmark also needs to be created for classifying the complexity of a sentence.

Table 1 Summary of the issues [22]

Sr. no.	Parameter	S.E.	NLP
1	Types of UML diagram generated	Yes	No
2	Generating textual information	Yes	No
3	Developing or using existing ontologies	Yes	Yes
4	Human factors	Yes	Yes
5	Level of noise in a sentence	Yes	Yes
6	Determining the complexity of sentence	No	Yes
7	Scanning of textual information for relevant information	Yes	Yes
8	Scanning of textual information for ambiguity	No	Yes

Scanning of textual information for relevant information

The input text must be scanned for getting the necessary information. This is done by making use of various NLP tools which are available for processing of text. It involves both semantic and syntactic processing of the text.

Scanning of textual information for ambiguity

This involves studying the input text for any ambiguous sentence and then removing those sentences [22] (Table 1).

4 Methodology for UML Diagram Generation

SE and NLP issues should be addressed before a good-quality UML diagram can be generated. Hence, we propose TextToUml (TTU) for the generation of UML diagrams [22]:

1. Define a **parameter** about the quality of N.L. text.
This involves classifying the text as in controlled language and uncontrolled language.
2. Understand the issues at level of **text** such as follows:
 - a. Level of noise,
 - b. Level of complexity of a sentence in terms of controlled language and uncontrolled language.
 - c. Determining the sentence in the following types:
 - i. Simple,
 - ii. Semi-complex,
 - iii. Complex.
3. Identify the **type** of diagram corresponding to the description given in the text.
4. Understand the **specification** of UML diagrams to be developed. Currently, there exist three different specifications:

- a. OMT,
 - b. Jacobson,
 - c. Booch.
5. Derive UML specification in tune with N.L. text available for all UML diagrams.
 6. Develop an interface between different ontologies and application to generate UML diagram.

We have already implemented an algorithm No_REM to remove the noise of an input text [23].

5 Results, Discussion, and Future Scope

The UML diagrams can be generated from natural language text [2–5]. However, the quality of the generated diagrams will depend upon how the issues in generating UML diagrams are dealt with. Based on these issues, UML diagrams can be generated from natural language text with high quality.

5.1 *Advantages of Current Work*

The analysis of issues in generation of UML diagrams from N.L. text has the following advantages:

- UML diagrams generated will be of good quality.
- Possibility of assessing the quality of N.L. text.
- Possibility of assessing the quality of documents generated from N.L. text.
- Help in better automation as two research areas are being addressed.
- A generic framework for addressing the interdisciplinary research can be developed.

5.2 *Disadvantages of Current Work*

The current work has the following disadvantages:

- Softwares for NLP are required to be downloaded and installed separately.
- Quality of N.L. text makes the UML diagrams from such a text, poor (in case sentence formation is not proper).
- Issues such as human factors (level of understanding of abstraction in UML diagrams) are not studied in-depth.

6 Application of Our Methodology in Non-software Engineering Environment

The technique can also be generalized to a non-software engineering environment. There has been work conducted on application of software engineering in various non-computing fields. For instance, UML diagrams can be drawn for those softwares which are utilized in building and construction fields of civil engineering.

7 Conclusion

In this paper, we have discussed the important issues while trying to create UML diagrams from natural language text. The natural language text forms the part of NLP and UML diagram is drawn in SE field. By the analysis in multiple disciplines, it may be possible to automate the creation of UML diagrams and also work toward achieving universal programmability [24]. The work also can be extended to check the quality of NL text, defining parameter for the quality of UML diagrams generated. By addressing the issues it will be possible to generate good-quality UML diagrams and their applicability at other domains can also be studied.

References

1. Pressman R.S., "Software Engineering A Practitioner's Approach", Tata McGraw Hill Publishing House, 7th addition. 2010.
2. Reynaldo Giganto, "Generating Class Models through Controlled Requirements", New Zealand Computer Science Research Conference (NZCSRSC) 2008, Apr 2008, Christchurch, New Zealand.
3. Sascha Konrad, Betty H.C. Cheng, "Automated Analysis of Natural Language Properties for UML Diagrams". In. Proc. International Conference in Computing Technologies. ICCT. pp no 2–5.
4. Priyanka More, Rashmi Phalnikar, "Generating UML Diagrams from Natural Language Specifications", International Journal of Applied Information Systems, Foundation of Computer Science, Vol-1-No-8, Apr-2012.
5. Mathias Landhauber, Sven J. Korner, Walter F. Tichy, "From Requirements to UML Models and Bac: how automatic processing of text can support requirements engineering", Software Quality General, March 2013, Vol 22, Issue 1, pp 121–149.
6. Stepane S. Some: "A Meta-Model for Textual Use Case Description", Journal of Object Technology (JOT), vol. 8, no. 7, November-December 2009, pages 87–106.
7. A. Cookburn, "Writing Effective Use Cases", Addison Wesley Publication, Page-No- 3–22, 2001.
8. Agt, H, "Supporting Software Language Engineering by Automated Domain Knowledge Acquisition", In. Proc. Kienzle, J. (ed.) MODELS 2011 Workshops. LNCS, vol. 7167, pp. 4–11. Springer, Heidelberg (2012).
9. Matthew Hause, "Finding Roles for Use-Cases", In. Proc. IEEE Information Professional June/July 2005, Pages 34–38.

10. Stephen L. Reed, Douglas B. Lenat, "Mapping Ontologies into Cyc", American Association of Artificial Intelligence, 2002.
11. Viliam Simko, Petr Kroha, Petr Hnetyinka, "Implemented domain model generation", Technical Report, Department of Distributed and Dependable Systems, Report No. D3S-TR-2013-03.
12. Sudha Verma, Sarah Vieweg, William J. Corvey, Leysia Palen1, James H. Martin1, Martha Palmer, Aaron Schram1 & Kenneth M. Anderson, "Natural Language Processing to the Rescue?: Extracting "Situational Awareness" Tweets During Mass Emergency", In. Proc. Association for the Advancement of Artificial Intelligence, 2011.
13. Artis Teilans, Yuri Merkurjev, Andris Grinbergs, "Simulation of UML models using ARENA", In. Proc. 6th WSEAS International Conference on SYSTEM SCIENCE and SIMULATION IN ENGINEERING, Venice, Italy, Nov 21–23, 2007. Pg No. 190–195.
14. Mathias Landhauber, Sven J. Korner, Walter F. Tichy, "From Requirements to UML Models and Back How Automatic Processing of Text Can Support Requirement Engineering", Springer's Software Quality Journal, March 2014, Vol. 22, Issue 1, Pages 121–149.
15. Imran Sarwar Bajwa, M. Abbas Choudhary, "Natural Language Processing Based automated system for UML Diagrams generation", In. Proc. 18th National Computer Conference 2006, Sandi Computer Society.
16. Imran S. Bajwa, Mark G. Lee, Behzad Bordbar, "SVBR Business Rules Generation from Natural Language Specification", In. Proc. Artificial Intelligence for Business Agility-Papers from AAAI 2011 Spring Symposium (SS-11-03), Pg. No. 2–8.
17. Imran S. Bajwa, Mark Lee, and Behzad Bordbar, "Resolving Syntactic Ambiguities in Natural Language Specification of Constraints", In. Proc. CILing 2012, Lecture Notes in Computer Science (LNCS) 7181, pp-178–187, 2012. Springer-Verlag, Heidelberg, 2012.
18. Imran S. Bajwa, M. Imran Siddique, M. Abbas Choudhary, "Rule Based Production Systems for Automatic Code Generation in Java", In. Proc. International Conference on Digital Information Management-ICDIM 2006, Pages 200–205, Pages 200–205 Bangalore, India.
19. Imran Sarwar Bajwa, M. Asif Naem, Ahsan Ali Chaudhri, Shahzad Ali, "A Controlled Natural Language Interface to Class Models", In. Proc. 13th International Conference on Enterprise Information Systems, Page No. 102–110. doi:[10.5220/0003509801020110](https://doi.org/10.5220/0003509801020110), Science and Technology Publications, SciTePress, Pakistan.
20. Imran Sarwar Bajwa, Irfan Hyder, "UCD-Generator A LESSA Application for Use Case Design", In. Proc. IEEE-International Conference on Information and Emerging Technologies, IEEE-ICIET, Pages 200–205 Karachi-Pakistan.
21. Dr. Prasanth Yalla, Nakul Sharma, "Combining Natural Language Processing and Software Engineering", In. Proc. International Conference in Recent Trends in Engineering Sciences (ICRTES-2014), Organized by Elsevier and IET, March 2014, Pg-No. 106–108, ISBN-978-93-5107-223-2.
22. Nakul Sharma, Dr. Prasanth Yalla, "Issues in developing UML diagrams from NL Text", In. Proc. WSEAS Recent Advances In Telecommunications, Informatics And Educational Technologies, Spain, ISBN: 978-1-61804-262-0, Pg-No. 139–145.
23. Dr. Prasanth Yalla, Nakul Sharma, "Parsing Natural Language Text of Use Case description", In. Proc. International Conference on IT in Business, Industry and Government (CSIBIG-2014), March 8–9, 2014. ISBN- 978-1-4789-3063-0.
24. Walter F. Tichy, Mathias Landhabuer, Sven J. Korner, "Universal Programmability- How AI Can Help?", In Proc. 2nd International Conference NFS sponsored workshop on Realizing Artificial Intelligence Synergies in Software Engineering, May 2013.
25. [Online] <http://rationalrose.com> retrieved on 04 January, 2012.
26. Oksana Nikiforova, Olegs Konstantins Gusarovs, Dace Ahilcenoka, Andrejs Bajovs, Luddmila Kozacenko, Nadezda Skindere, Dainis Ungurs, "Development of BRAINTOOL from the two-hemisphere model based on the two-hemisphere Model transformation itself", In. Proc International Conference on Applied Information and Communication Technologies (AICT 2013), 25–26 April 2013, Latvin. Page 267–274.

2-D Photonic Crystal-Based Solar Cell

Mehra Rekha, Mahnot Neha and Maheshwary Shikha

Abstract Light trapping inside a solar cell is a very important parameter needed to be studied at the time of its designing. All conventional silicon solar cells which are currently in use have low light trapping, in turn providing low efficiency. In this paper, null radius defects are introduced in photonic crystals to improve the light trapping capacity of the solar cell. The paper deals with design of photonic crystals with null radius defect and its use in solar cell to increase its efficiency. In this proposed research work, power spectrum of solar cell has been studied and absorptions of photonic crystal-based solar cell and conventional silicon solar cell are compared at different input wavelengths.

Keywords Photonic crystal • Solar cell • Null radius defect • Silicon • Light trapping • Absorption

1 Introduction

Today, there is a need to replace non-renewable sources of energy. Solar energy is a vast source of energy. It, in fact, has the potential to replace conventional source of energy to mankind. The energy delivered by the Sun in 1 h is enough to be used by people in a whole year. Current solar cell systems operate at less than 30 % power conversion efficiency, but the theoretical limit is greater than 86 % [1]. The reason for limited efficiency is the fact that all the photons do not generate electron–hole pair. The first reason behind this is the refractive index contrast of air and solar cell material; some photons are reflected. The second reason is its low absorption

Mehra Rekha (✉) · Mahnot Neha · Maheshwary Shikha
Government Engineering College, Ajmer, India
e-mail: mehrarekha710@gmail.com

Mahnot Neha
e-mail: neha29mahnot@gmail.com

Maheshwary Shikha
e-mail: shikhasm1992@gmail.com

coefficient, i.e. all the photons do not get absorbed and leave the solar cell system. The third reason is that system itself emits radiation. So it is needed to assure that photons enter the solar cell device or it should prevent photons from leaving. These purposes can be termed as 'light trapping'. Light trapping in a solar cell allows either to increase its efficiency by enlarging the number of used photons, or it allows decreasing the solar cell's thickness which is the basic requirement of thin film solar cell.

1-D and 2-D photonic crystals are the centre of attraction for researchers these days. 1-D photonic crystal (i.e. grating) can be used to reduce reflection as anti-reflective coating [2] or can be used as back reflector. 2-D photonic crystal can also be used as back reflector [3]. Double-layer anti-reflective coating is also used to reduce reflection. Materials such as SiO_2 , Si_3N_4 and TiO_2 can be used as ARC [4]. Different photonic crystal arrangements can also be used to trap light, i.e. to increase absorption inside the solar cell. To increase the absorption in solar cell further, different defects are being created inside the photonic crystal [5]. Here in this work null radius defects are created to increase absorption in solar cell.

2 Photonic Crystal

Photonic crystals are periodic micro- or nanostructures that affect the motion of photons exactly in the same way as ionic lattice affects the electrons [6]. Photonic crystals are periodic repetition of lower and higher dielectric constants. The phenomenon occurs when the period of photonic crystal (hole radius and a hole to hole spacing) is less than the wavelength of the light. Whether photons propagate through these structures or not depends on their wavelength. The wavelengths which are allowed to travel through these structures are known as modes, and group of modes is called as bands. The bands which are not allowed to propagate through these structures form photonic band gaps. Some photons of wavelength within the band gap are prohibited from propagation in one, or all the direction inside a photonic crystal, providing the possibility to confine and trap the light in a cage. Hence, these photonic crystals can be used in solar cell to trap the photons.

Photonic crystals are classified as 1-D, 2-D and 3-D. They have periodicity, i.e. alternate layers of lower and higher refractive index in one, two and three dimensions, respectively. Bragg grating is the example of one-dimensional photonic crystal; photonic crystal fibre and opal are, respectively, two- and three-dimensional photonic crystals (Fig. 1).

3 Null Radius Defect

First, the defect-free photonic crystal slab is investigated for thin film solar cells with rectangular lattices, and then sub-lattices of defects are introduced to further enhance absorption.

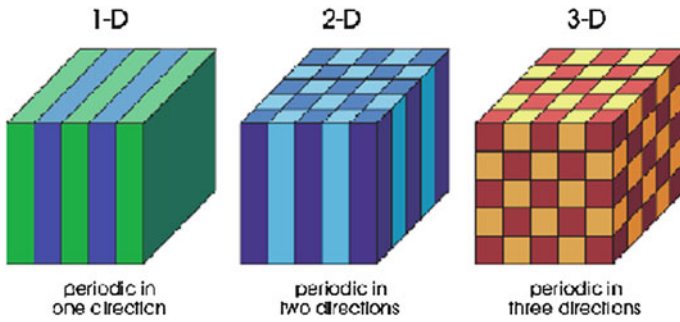


Fig. 1 1-D, 2-D and 3-D photonic crystals are represented, respectively. Different colours in cube denote materials with different refractive indexes [6] (color figure online)

In this design, a 2-D photonic crystal having rectangular lattice is used. Air holes of diameter 350 nm ($r = 175$ nm) and hole to hole spacing (a) of 500 nm are used. Null radius defect has been created in this photonic crystal to further increase the trapping of photons.

Null radius defect is created in photonic crystal by reducing the radius of some particular air holes to null (no air holes at the particular positions). Here in this design null radius defects are created at all these points: (0,1,1), (0,1,3), (0,1,5), (0,1,7), (0,3,2), (0,3,4), (0,3,6), (0,5,1), (0,5,3), (0,5,5), (0,5,7), (0,7,2), (0,7,4) and (0,7,6). Figure 2 shows the basic view of photonic crystal structure for defect-free photonic crystal and null radius defect in photonic crystal.

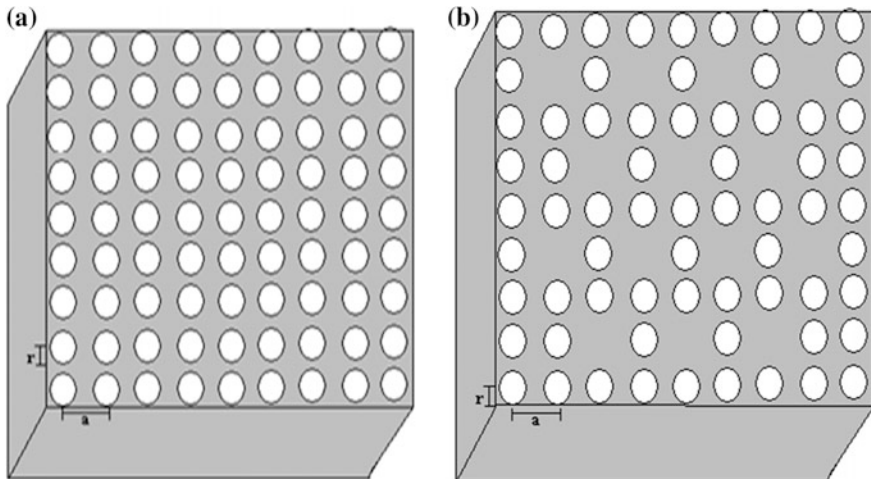


Fig. 2 **a** Defect-free photonic crystal layout **b** null radius defect in photonic crystal layout. Figure shows x - y plane only

4 Design and Simulation

The basic design consists of two layers of anti-reflective coating on photonic crystal and a back reflector. SiO_2 and Si_3N_4 (with refractive index 1.5 and 2.016, respectively) are used as an anti-reflective coating and silver is used as the back reflector. The layer of anti-reflective coating is used to reduce reflection of photons at the surface so that more photons can enter in the solar cell. Back reflector is used to reflect the unused photons back to photonic crystal (Fig. 3).

The layout of the design is drawn on Opti FDTD (Finite Difference Time Domain). To create a defined material profile in FDTD, we just need the refractive index of material. AMPL boundary condition is used with source wavelength in the range of 400–800 nm. This range is chosen so, because AM. 1.5 G spectrum [7] has the maximum solar irradiance in this region.

Figures 4 and 5 show the reflectance and transmittance of photonic crystal with introduced null radius defects. The absorption for the given design can be calculated using the following formula:

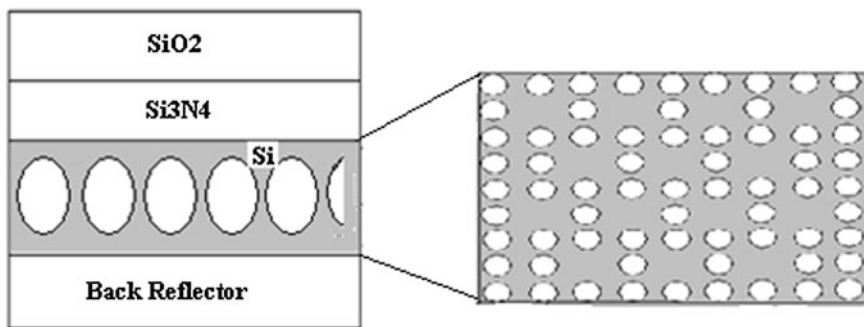


Fig. 3 Basic design of solar cell based on photonic crystal with null radius defect

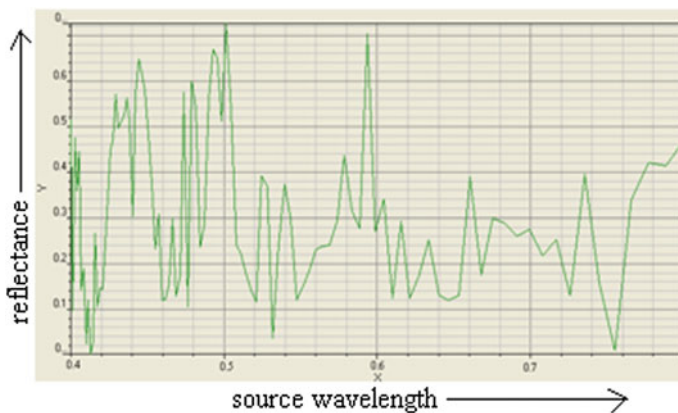


Fig. 4 Reflectance at different input wavelengths at $a = 500$ nm and $r = 175$ nm

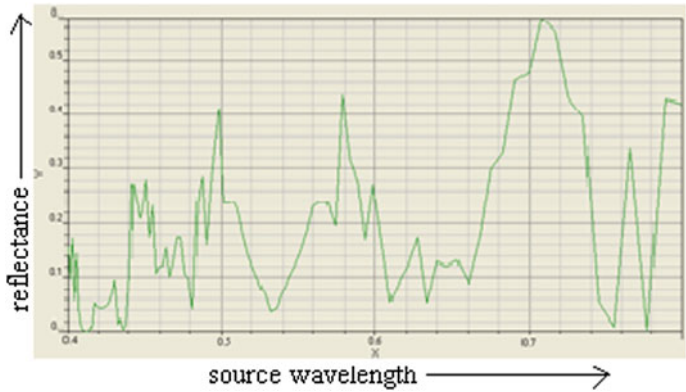
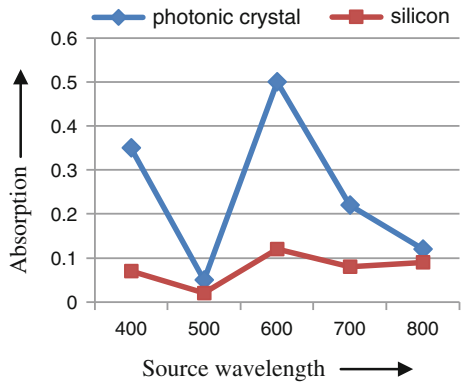


Fig. 5 Transmittance at different input wavelengths at $a = 500\text{ nm}$ and $r = 175\text{ nm}$

Table 1 Comparison of absorption of null radius defect in photonic crystal solar cell and conventional silicon solar cell

Source wavelength	Absorption of photonic crystal with null radius defects solar cell	Absorption of silicon-based semiconductor solar cell
400	0.35	0.07
500	0.05	0.02
600	0.5	0.12
700	0.22	0.08
800	0.12	0.09

Fig. 6 Absorption at different input wavelengths at $a = 500\text{ nm}$ and $r = 175\text{ nm}$



$$A = 1 - R - T. \tag{1}$$

In this work, the absorption of photonic crystal solar cell with null radius defect and the semiconductor solar cell has been compared. The power spectrum of

photonic crystal solar cell with null radius is found to be far better than that of conventional solar cell. The graph has been plotted for the input range of 400–800 nm in Fig. 6 (Table 1).

5 Conclusion

We have presented periodically textured Si to increase light trapping inside a solar cell. We have designed null radius defects in 2-D photonic crystal formed by etching holes in silicon wafer. We have compared the normalized absorbed power of photonic crystal solar cell having null radius defect to that of conventional silicon solar cell, whereas in [5] current densities at different defect diameters are presented.

In this paper, absorption is taken as key parameter because absorption can be easily linked to the reflectance and transmittance in solar cell. Light trapping can be further enhanced by either reducing reflectance at front end or by increasing reflectance at rear end. This can be done using anti-reflective coating [5] or by using texturing at front surface. It can also be linked to the techniques like use of Bragg grating [8] or diffraction grating [9] to direct photons back into solar cell. So the use of absorption in the analysis of 2-D photonic crystal is an effective approach as compared to that of current density [5] for light trapping in a solar cell as predicted.

References

1. Munday Jeremy: Photonic structured solar cells to enhance current and voltage. SPIE Newsroom. doi:[10.1117/2.1201405.005471](https://doi.org/10.1117/2.1201405.005471), 19 May 2014.
2. Buencuerpo Jeronimo, Munioz-Camuniez Luis E., Maria L. Dotor, Pablo A. Postigo: Optical absorption enhancement in a hybrid system photonic crystal – thin substrate for photovoltaic applications. OPTICS EXPRESS, A452, Vol. 20, No. S4, #160029, 2 July 2012.
3. Gupta Nikhil Deep, Janyani Vijay: Efficiency Enhancement in the thin film GaAs solar cell using Photonic Crystal as a back reflector. WRAP'13, New Dehi, India, IEEE Dec.17–18, 2013.
4. Ali Khuram, Sohail A. Khan, Jafri M. Z. Mat: Effect of Double Layer (SiO₂/TiO₂) Anti-reflective Coating on Silicon Solar Cells. Electrochem. Sci., 9 (2014) 7865–7874, 28 October 2014.
5. Kelsey A. Whitesell, Dennis M. Callahan, Harry Atwater: Silicon Solar Cell Light-Trapping Using Defect Mode PhotonicCrystals. SPIE Vol. 8620, 86200D © 2013 SPIECCCode: 0277-786X/13/\$18 doi:[10.1117/12.2005450](https://doi.org/10.1117/12.2005450).
6. Russell P. St. J. and Dettmer R.: A neat idea photonic crystal fibre. IEEE Review, vol. 47, pp. 19–23, Sept. 2001.
7. Gueymard: The sun's total and spectral irradiance for solar energy applications and solar radiation models. Solar Energy, Volume 76, Issue 4, April 2004, Pages 423–453.

8. Peter Bermel, Chiyan Luo, Lirong Zeng, Lionel C. Kimerling, and John D. Joannopoulos. "Improving thin-film crystalline silicon solar cell efficiencies with photonic crystals", *OPTICS EXPRESS* 16986, Vol. 15, No. 25, December 2007.
9. James G. Mutitu, Shouyuan Shi, Allen Barnett, and Dennis W. Prather, "Angular Selective Light Filter Based on Photonic Crystals for Photovoltaic Applications" *IEEE Photonics Journal*, Volume 2, Number 3, 490–499, June 2010.

Parallel Implantation of Frequent Itemset Mining Using Inverted Matrix Based on OpenCL

Pratipalsinh Zala, Hiren Kotadiya and Sanjay Bhanderi

Abstract Extracting knowledge in the form of frequent itemsets and association rules deserves great importance in the field of data mining. Apriori algorithm suffers from multiple scans of the database and thus forms high memory dependency. On the other hand frequent pattern tree (FP tree) growth algorithm becomes impractical for large databases due to memory-based data structure. An efficient approach of inverted matrix with COFI (co-occurrence frequent item) tree alleviates disadvantages of both the above-mentioned algorithms. For massively large computations, modern GPUs provide a large set of parallel processors which facilitate in general-purpose computing. General purpose graphical processing unit (GPGPU) is way of utilizing the existing GPU for general purpose use. Open computing language (OpenCL) provides a standard for cross-platform programming on modern processors such as many-core CPUs and GPUs. As inverted matrix approach is advantageous over other algorithms, it is desirable to form it parallel to OpenCL. We have proposed a new technique called CLInverted matrix itemset mining, which is an advancement over existing techniques and contributes to load sharing. The proposed architecture in this paper highlights the inverted matrix approach implantation based on OpenCL framework. In experiments we have compared the results of serial and parallel versions of the proposed approach on various OpenCL devices.

Keywords Frequent itemset • Opencl • GPGPU • Inverted matrix • COFI tree

Pratipalsinh Zala (✉) • Hiren Kotadiya • Sanjay Bhanderi
Faculty of PG Studies-MEF Group of Institutions,
Department of Computer Engineering, Rajkot 360003, India
e-mail: pratipalzala16@gmail.com

Hiren Kotadiya
e-mail: hirenkotadiya@gmail.com

Sanjay Bhanderi
e-mail: sanjay.bhanderi@marwadieducation.edu.in

1 Introduction

Data mining is considered as discovery of knowledge which is useful, from a huge amount of data. Association rule mining is one of the important aspects of data mining. Association rule mining can be defined as finding association between itemsets or items in the database. These kinds of association rules can be helpful for finding the customer buying habits, e.g., market basket analysis. Existing solutions for identifying frequent patterns, sequential or parallel, suffer from many obstacles as high memory dependency due to multiple scan of database, huge memory requirements, etc. Modern processor architectures encompass ability of parallelism as a way to performance improvement. GPGPU provides the facility of utilizing the graphics processing unit (GPU) for general-purpose applications. OpenCL is an open-source standard which gives software developers, portable and efficient access to the power of CPUs, GPUs, and other processors for general-purpose parallel computing [1, 2]. There exist solutions of parallel implementation of AES cryptography algorithm. AES implantation based on OpenCL [3] determines the efficiency over sequential algorithm of advanced encryption standard algorithm.

OpenCL platform model contains multiple numbers of compute devices which in turn contain multiple compute units. Processing elements are the basic units of compute unit. Processing elements can be considered as threads of execution. Host (which is CPU) assigns the set of similar instructions to compute units. Multiple instruction set execution environment of OpenCL platform forms parallel environment on GPU.

The proposed architecture of this paper shows the parallel formation of approach of inverted matrix with COFI (co-occurrence frequent item) tree [4] mining. The working of this approach is divided into two phases as (1) construction of inverted and (2) building and mining COFI tree.

2 Related Work

Let $I = \{i_1, i_2, i_3, \dots, i_n\}$ be a set of items and $T = \{t_1, t_2, t_3, \dots, t_m\}$ be a set of transactions. D is a database which consists set of transactions. Each transaction $t_a (a = 1, 2, 3, \dots, m)$ belongs to T containing a subset of items $i_b (b = 1, 2, 3, \dots, n)$ belonging to I . Therefore, all transactions in T are subsets of the item set I . Any itemset I_f is said to be frequent if its support count is identical to or greater than a given minimum support threshold. Frequent itemset I_f is called a frequent m -itemset, if it contains m items. 1-itemset is also called frequent item [5, 6]. Basically frequent itemset mining algorithm gives set of all size frequent itemsets.

As stated above, Apriori and FP tree [7] algorithms suffer from various issues. In dynamic hashing and pruning (DHP) [8], candidate space is reduced through pre-calculating proximate support for $m + 1$ itemset while counting m -itemset by constructing a hash table. In DHP, also transactions are removed which do not

contain any frequent items, which is called transaction trimming. Both properties of pruning and trimming become obstacles and hence make DHP unreal in many cases. MLFPT (multiple local frequent pattern tree) [9] is parallel frequent pattern mining algorithm. As this approach is based on FP tree, it only requires two full dataset scans. Inverted matrix approach [10] is one of the efficient approaches for frequent itemset mining. Once the inverted matrix is constructed (which requires only two scans of the original database), repetitive scanning of the database for frequent itemsets of different support thresholds is avoided. On the other hand, COFI trees which are later constructed and mined are comparatively smaller data structures than FP tree. In this way abundant memory requirement problem is obviated. The following three steps describe the process of frequent itemset mining by inverted matrix approach:

2.1 Creation of Inverted Matrix

Inverted matrix builds a tabular structure which consists of pointer in each field. This table is formed from the original transactional database. Each pointer points to the next item in the transaction. The first column of table is formed by arranging items according to ascending order of their frequency values in the transactions. Construction of this table needs only two scans of the original database. One big advantage of this data structure is it avoids superfluous processing.

2.2 Building COFI Trees

COFI tree is much more similar to FP tree except that it contains bidirectional links. In the COFI tree, child node is as frequent as or more frequent than the parent node. COFI trees are small trees compared to FP tree, so they require less memory space. For each and every item of the inverted matrix, individual COFI trees are built. If any predefined support threshold is given, then COFI tree for the items in inverted matrix which do not satisfy that support is not built. COFI tree of most frequent items in inverted matrix is not generated. Bidirectional pointers help in procedure of mining COFI trees in top-down and bottom-up traversal.

2.3 Mining COFI Trees

Generation of COFI trees is a stepwise process. In the traditional way COFI trees for all frequent items are not generated together. Individual COFI tree is constructed, mined, and discarded before COFI trees for the next items are constructed.

With the use of support count and participation count from all branches of tree, candidate frequent itemsets are determined and stored temporarily in the list. At the end, when all branches are processed infrequent itemsets are eliminated.

3 Parallel Inverted Matrix Approach Using OpenCL Framework

In this paper, architecture and algorithm of parallel inverted matrix approach in OpenCL framework are presented. As shown in Fig. 1, the architecture is divided into two phases. The first phase deals with construction of inverted matrix and the second phase performs work of building and mining COFI trees. In the first phase transactional database is given as input to host. The host is considered as the CPU of the system. Environment of GPU is considered as a collection of work groups. Again, these work groups are formed by many work items. Work items are considered as threads or computational units. All these threads have their own private memory. The memory of each work group is shared among its work items. Work items of the same work group can communicate and share data among them. Transactions in database are equally distributed among available work groups. The first scan process to build inverted matrix takes place on each work group. On each work group a subset of transactions are scanned and unique items in those transactions with their frequency counts are found. Counting global frequency of all items is done by broadcasting these local frequencies. The process of inverted matrix building starts by collecting items with their frequency values from all work groups and are completed on master node by performing a second scan on the original database. In this way inverted matrix is built in parallel as part of the future work of [10].

The second phase performs COFI tree building and mining. Inverted matrix is replicated over the existing work groups in their respective memory. This process of assigning inverted matrix to all work groups reduces communication overhead among the work groups. In Fig. 1, X work groups are available. There are m unique items in the inverted matrix (assume that $m > X$). As it is clear from inverted matrix, the most frequent item is located at the highest index position. Items on the first and second last positions are assigned to work group 1. The next indexed positioned item and third last positioned item are assigned to work group 2. After assigning items on X th work-group, the next consecutive pair of items are assigned to work group 1 again until all frequent items are distributed among X work groups. In this way all frequent items are grouped into X itemsets and are distributed among X work groups. All the frequent itemsets discovered on every work group are gathered back on the master node. Work items in the work groups share the task assigned to that particular work group. Generally, it is observed that the depth of COFI tree of an item which is least frequent is the highest. It has been observed that there is high possibility that COFI tree for the less frequent item requires more time

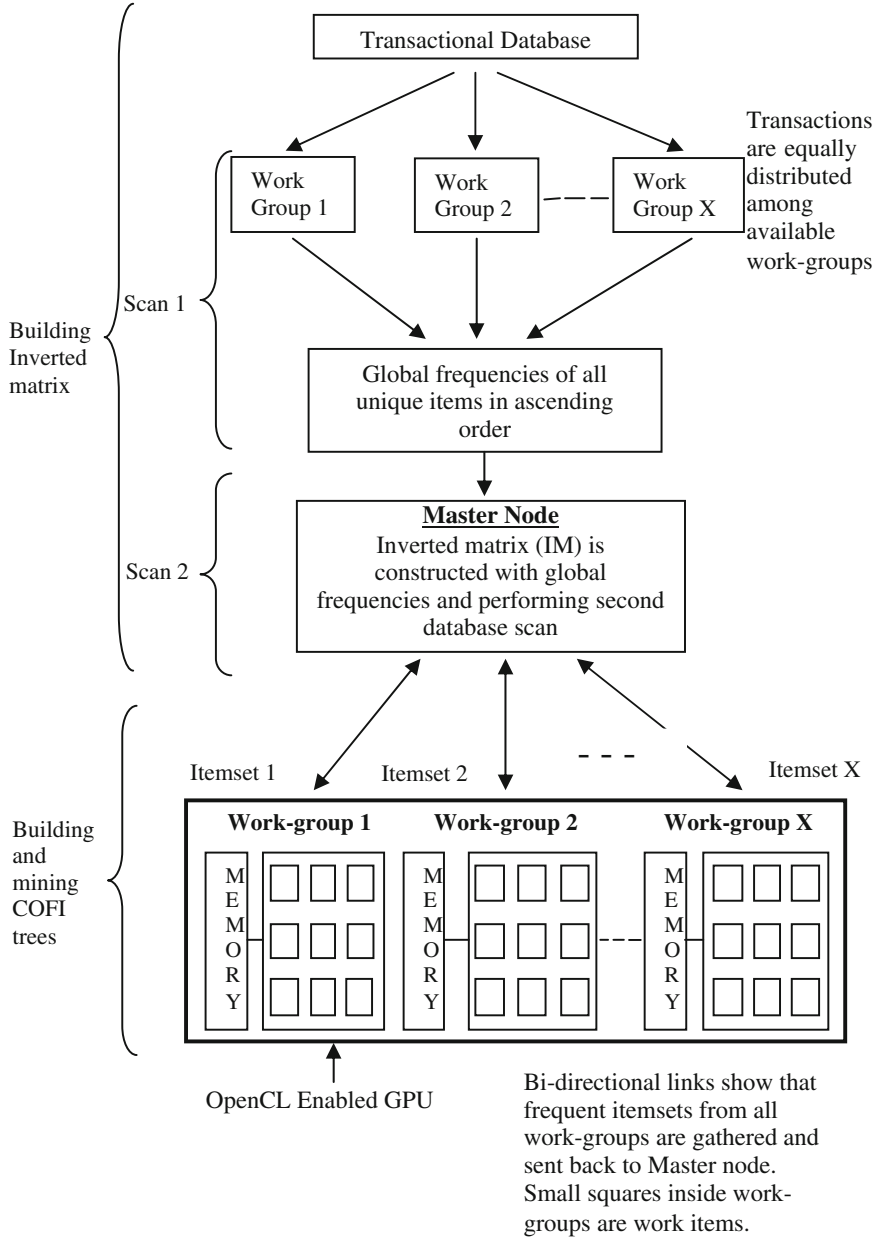


Fig. 1 Implantation of inverted matrix approach on OpenCL framework

Algorithm: CLInverted matrix itemset mining

Input: Transactional database with transaction set $T = \{T_1, T_2, \dots, T_m\}$ and unique item set $I = \{I_1, I_2, \dots, I_n\}$.

Output: Frequent itemsets

/* Creating Inverted matrix */

1. Equally distribute transactions among available work-groups $W = \{WG_1, WG_2, \dots, WG_r\}$ (assuming $m > r$).
2. In first scan of all subsets of transactions allotted to each work-group, unique items with their frequency counts are discovered.
3. Generate global frequency for all the items by broadcasting local frequencies obtained in step 2.
4. Master node performs second scan on transactional database and generate Inverted matrix.

/* Building and mining COFI trees */

5. Distribute N frequent items of Inverted matrix amongst r work-groups (assuming $N > r$) as following:
 - 5.1. Assign least frequent item (first item in IM) and second most frequent item (as COFI tree for most frequent item is not generated) to WG_1 , next least frequent item and third most frequent item to WG_2 and so on up to r^{th} work-group. Assign $(r+1)^{th}$ least frequent item and $(N-r-2)^{th}$ most frequent item to WG_1 again until all frequent items are distributed amongst r work-groups.
 - 5.2. Build COFI trees for all items assigned to each work-group.
 - 5.3. Mine all COFI trees built on every work-group.
 - 5.4. Send all mined frequent itemsets on each WG back to Master node.

Fig. 2 Algorithm of CLInverted matrix itemset mining

to mine compared to the COFI tree for the more frequent item. In this way, load balancing is achieved by the proposed technique. Figure 1 presents the architecture of the proposed work.

Alternate loop splitting (ALS) and block loop splitting (BLS) [9] are existing techniques for load balancing. Experimental results of [9] show that while considering load balancing in frequent pattern mining using inverted matrix, ALS shows better results than BLS. The proposed algorithm for CLInverted matrix itemset mining is shown in Fig. 2.

4 Implementation and Results

In our work we have implemented serial as well as parallel inverted matrix on three different OpenCL devices and measured results accordingly. As input to the approach we have taken mushroom dataset [11] with around 8000 transactions, 120

unique items, and with an average of 23 items in a transaction. We have implemented parallel as well as serial inverted matrix algorithm in Visual studio 2010 Express Edition with C++. Following are the OpenCL devices on which we implanted our approach.

1. Intel(R) Core(TM) i3-3227U CPU@1.90 GHz with 4 GB RAM.
2. AMD radeon HD 8670 M with memory size 1024 (in MB) (Intel(R) core i3 CPU with 4 GB RAM)
3. ATI mobility radeon HD5870 (5000 series) with memory size 1024 (in MB) (Intel(R) dual core CPU with 4 GB RAM)

For each device mentioned above, the information in brackets shows CPU details of the corresponding devices. We have measured the results for different number of transactions in the input dataset. In the results presented in graphs, the X-axis represents runtime of an algorithm in milliseconds and the Y-axis represents different OpenCL devices with their serial and parallel utilization.

From the results shown in Fig. 3, it was noticed for all devices that parallel runtime is more than that of serial runtime in the proposed approach. This signifies that for parallel version of algorithm input size should be large enough to overcome drawbacks of parallel algorithms as task distribution and communication overhead. Because of such issues, we found more runtime taken by parallel approach than the serial one. As we doubled the size of input dataset in terms of number of transactions as shown in Fig. 4, betterment in results of parallel approach on OpenCL devices with GPUs was discovered. In a device without GPU, the result was negligibly improved. Finally, when we increased dataset size to around 25,000 transactions, major difference in runtime of parallel approach was gained, compared to serial version of approach. In this way more improvement is possible even for large size of datasets (Fig. 5).

Fig. 3 Result comparison for number of transactions ~ 8000

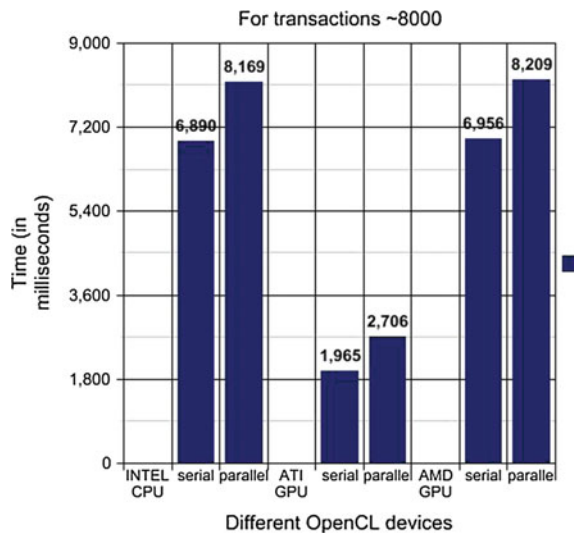


Fig. 4 Result comparison for number of transactions ~17,000

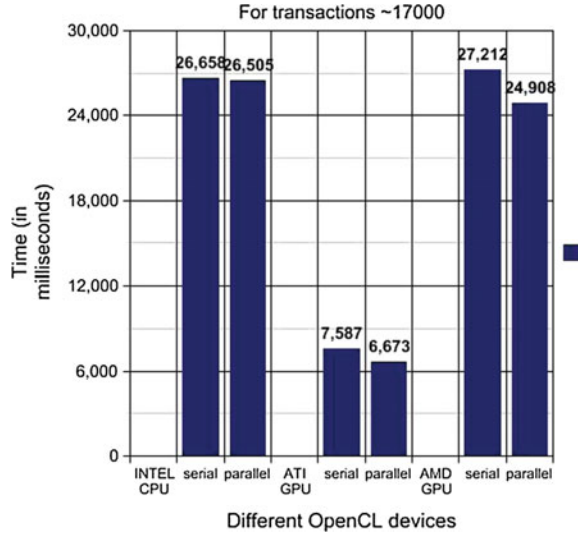
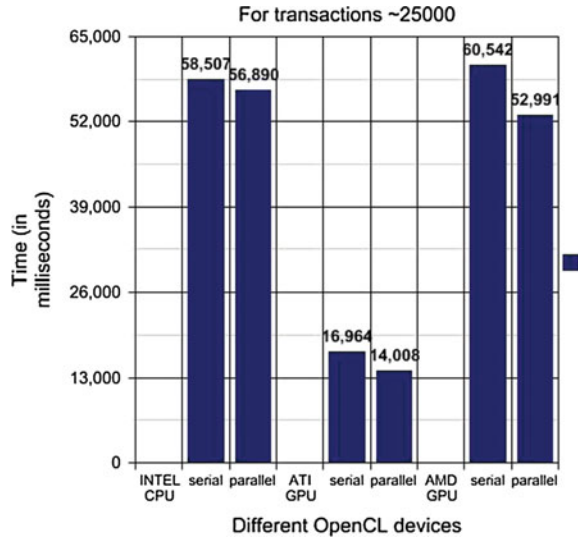


Fig. 5 Result comparison for number of transactions ~25,000



5 Conclusion and Future Work

Parallel inverted matrix approach is novel and far more advantageous over traditional approaches as concerns frequent itemset mining. GPGPU concept along with OpenCL has extended the capability of routine system GPUs for general-purpose parallel applications. We use inverted matrix approach for better utilization of OpenCL devices to form a parallel version of it.

As mentioned earlier, we parallel implemented scan I of the original dataset to find the frequency of all the unique items. After that we had to serially implement the inverted matrix creation as in this phase pointer pointing to the next item is dependent on the previous pointers. New data structure as replacement of inverted matrix or change in the structure of inverted matrix is suggested for purpose of parallel formation. Also, this work can be extended with datasets of different nature and size.

References

1. Tompson, J., Schlachter, K.: An Introduction to the OpenCL Programming Model. Person Education (2012).
2. Khronos group, <http://www.khronos.org/opencl>.
3. Gervasi, O., Russo, D., Vella, F.: The AES Implantation Based on OpenCL for Multi/Many Core Architecture. In: IEEE International Conference on Computational Science and Its Applications (ICCSA), pp. 129–134, IEEE (2010).
4. El-Hajj, M., Zaïane, O. R.: COFI-tree Mining: A New Approach to Pattern Growth with Reduced Candidacy Generation. In: Workshop on Frequent Itemset Mining Implementations (FIMI'03) in Conjunction with IEEE-ICDM, (2003).
5. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In: Proceedings 20th International Conference on Very Large Databases (VLDB), vol. 1215, pp. 487–499, ACM (1994).
6. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd ed., Morgan Kaufmann Publisher, San Francisco (2006).
7. Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation. In: ACM SIGMOD Record, vol. 29, no. 2, pp. 1–12, ACM (2000).
8. Park, J. S., Chen, M., Yu, P. S.: An Effective Hash Based Algorithm for Mining Association Rules. In Proceedings of ACM SIGMOD Conference, pp. 175–186, ACM Press, New York (1995).
9. Zaïane, O. R., El-Hajj, M., Lu, P.: Fast Parallel Association Rule Mining without Candidacy Generation. In: Proceedings IEEE International Conference on Data Mining (ICDM), pp. 665–668, IEEE (2001).
10. El-Hajj, M., Zaïane, O. R.: Inverted Matrix: Efficient Discovery of Frequent Items in Large Datasets in the Context of Interactive Mining. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 109–118, ACM (2003).
11. Frequent Itemset Mining Dataset Repository, <http://fimi.ua.ac.be/data/mushroom.dat>.
12. El-Hajj, M., Zaïane, O. R.: Parallel Association Rule Mining with Minimum Inter-processor Communication. In: Proceedings of 14th International Workshop on Database and Expert Systems Applications, pp. 519–523, IEEE (2003).

Extended Visual Secret Sharing with Cover Images Using Halftoning

Abhishek Mishra and Ashutosh Gupta

Abstract An extended visual cryptography scheme (EVCS) is a category of visual cryptography scheme (VCS) in which secret image is encoded into multiple shares of meaningful images. It has two additional images which are covering shares by the end of the encoding process. These meaningful shares are created by different approaches. The purpose of cover images (meaningful images) is to hide the secret image under it. In this paper, we propose an extended visual cryptography scheme with cover images using halftoning method. The halftoning method we designed for conversion of gray level image into binary image is based on dithering. The obtained halftoned image is transformed into multiple shares that are distributed to the participants. These shares are finally covered with some cover images to obtain meaningful shares. The experimental results and analysis show that the proposed scheme has satisfactory results.

Keywords Extended visual cryptography · Halftoning · Cover images · Security

1 Introduction

There is huge increase in transmission of data over network for instant access or distribution of data. As data are available in many forms that include text, image, audio, and video, images are one of the important data items. Today, researchers use visual cryptography schemes for distribution of secret images in the form of share (or shadow) images. The concept of secret sharing was first introduced by Naor and Shamir [1] and it is one of the active research areas in information security. There are a variety of ways through which information can be secure

Abhishek Mishra (✉)
IFTM University, Moradabad, India
e-mail: abhimishra2@gmail.com

Ashutosh Gupta
MJP Rohilkhand University, Bareilly, India
e-mail: ashutosh3333@gmail.com

including image hiding, watermarking, key exchange, authentication etc. However, these methods have a drawback that secret image is concealed in a single information carrier. If this concealed information is lost, there is no way to retrieve it.

Such problem can be overcome by visual secret sharing (VCS) scheme introduced by Naor and Shamir [1–3]. The scheme splits a secret image into multiple parts, also called share or shadow images and distributes each share among the number of participants. A subset of participants can only reveal the secret image by stacking the shares in some predefined order.

An extended visual cryptography scheme (EVCS) is a type of VCS in which secret image is encoded into multiple shares of meaningful images. These meaningful shares are created by different approaches. The purpose of cover images (meaningful images) is to hide the secret image under it. It is also a practical fact that secret images are not always in the form of monochrome. They may be in the form of color or gray level images. The same explanation also holds for cover images. This necessitates that there should be some transformation mechanism that converts the color or gray level images into monochrome images. The most common transformation to convert color or gray level image into binary image is halftoning. In this paper, we propose a visual cryptography scheme with cover images using halftoning method. The halftoning method we designed for conversion of gray level image into binary image is based on dithering. The rest of the paper is organized as follows: Sect. 2 explains the basics of extended VCS and common halftoning techniques. In Sect. 3, we describe our proposed EVCS scheme followed by experimental results in Sect. 4. Finally, Sect. 5 concludes the work.

2 Background and Related Work

Ateniese et al. [4] first proposed the concept of extended visual cryptography scheme (EVCS), which is a special category of VCS where secret image is encoded into multiple meaningful shares. An extended VCS requires more inputs compared to traditional VCS and these additional inputs are also images that work as cover shares (or images) after completion of the encoding process.

In the first step, the shares of a secret image are generated in the usual way by applying a VCS scheme. The first step is common for all varieties of images. Many visual cryptography schemes have been developed in the recent past [5–9]. The next task is how these generated shares are embedded or hidden within the chosen meaningful (cover) images. This requires a suitable method so that pixels from the secret shares remain distinct over the chosen cover image. The input image may be monochrome or grayscale or colored in nature. The same argument is also applied for chosen cover images. In this case, it is essential to convert the secret image into a monochrome image as the traditional or newly proposed secret sharing scheme [10, 11] relies on binary images before these shares are hidden within meaningful shares.

The conversion of secret image into binary image is performed through halftoning. Halftone visual cryptography (HVC) proposed by Wang et al. [12] adds digital halftoning techniques to extend the area of visual cryptography. The figure shows the dithering matrix of the gray-levels 0–9 to obtain proper halftoned patterns. Specifically, in VSS schemes, meaningful visual information can be used to encode a secret image into halftone shares. Halftoning is a method that simulates the grayscale of pixels by utilizing the density of printed dots. The human visual system can record only the overall intensity and integrate the fine detail in an image viewed from a distance. The denser the dot, the darker the image; in contrast, the sparser the dots, the lighter the image. Thus, one can use either black or white colors to simulate a continuous tone such that continuous-tone image can be changed to binary image.

For example, Fig. 1a shows a gray-level image that is transformed into a binary image [13] shown in Fig. 1b with black and white dots using halftoning. However, Fig. 1b is a binary image, and the human visual system can still perceive the gray level changes as it is a gray level image. Mostly, the visual cryptographic methods are designed for binary images, so the halftoning method is used to convert a gray-level image into a binary image. Thus, one can use the Naor and Shamir (2, 2)-threshold VSS scheme to encrypt Fig. 3b. The result is shown in Fig. 3c, which demonstrates the applicability and feasibility of using halftoning to construct a VSS scheme for gray-level images. The above arguments prove that halftoning techniques are useful preprocessing steps in visual cryptography to convert grayscale images to binary images. Since, halftoning reduces the quality of an image when applied on grayscale image and further degradation in image quality is due to VSS schemes, thus the overall effect is moderate degradation in image quality. This becomes an important parameter in a visual cryptography scheme along with some other issues as image expansion [14] and conciliation of the security of the scheme [15].

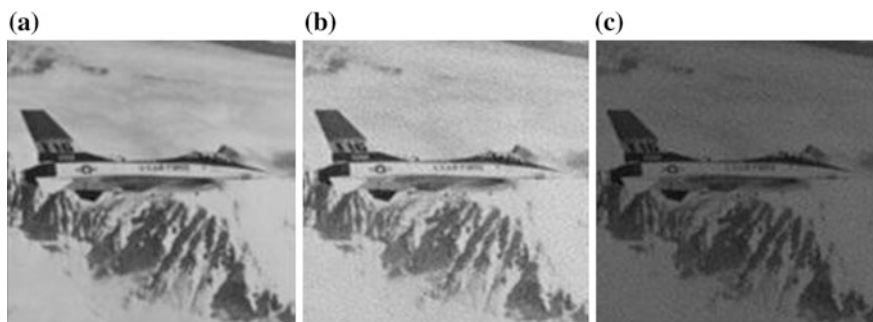


Fig. 1 Halftoning (512×512 pixels). **a** A continuous-tone image. **b** A halftone image. **c** The stacked image (1024×1024)

3 Proposed Scheme

In this section, we propose a halftoning method and visual secret sharing where shares are generated in some meaningful form. The meaningful shares are generated by hiding the secret shares through some cover images.

3.1 Halftone an Image

This section describes a halftone method where gray-level image is converted into a binary image. The method used for halftone conversion is ordered dithering. Let I and P be an $m \times m$ gray level and corresponding halftoned binary image respectively. Let $M(m, m)$ be the random image consisting of $L = 16$ gray levels. The mapping of $I(i, j)$ to $P(i, j)$ is done by normalizing the gray level value of $I(i, j)$ and $M(i, j)$. The proposed halftoned algorithm computes and compares the normalized value of $I(i, j)$ with $M(i, j)$. If normalized value of $I(i, j)$ is greater than or equal to normalized value of $M(i, j)$, then pixel value of halftone image is set to 1, otherwise 0. The algorithm for converting grayscale image to halftoned image is shown in Fig. 2.

3.2 Share Generation

This section presents the method for generating the shares from the halftone image P obtained from algorithm 1 with some meaningful image (also called cover image) taken as inputs. Next are the n random binary matrices M_k where $1 \leq k \leq n$ is generated. The n intermediate matrices I_k are generated by applying XOR operation according to the following rule:

Algorithm 1: Algorithm for constructing Halftone image

Pre-condition: A gray scale images I and M of size $N \times N$ and $L=16$
 Post-condition: Halftone image P .

```

for i = 1 to N
  for j = 1 to N
    if  $I(i, j)/256 \geq M(i, j)/L$ 
       $P(i, j) = 1$ ;
    else
       $P(i, j) = 0$ ;
    endif
  end for
end for

```

Fig. 2 Algorithm for halftone image

1. If $P(i, j) = 0$ (i.e., black pixel), then $I_k = M_n \oplus M_k$. (a) If values of both M_n and M_k are 0 or 1 respectively, then $I_k = 0$. (b) If values of M_n and M_k are either 0 or 1, then $I_k = 1$. This implies that there is 50 % probability that black pixel passes as it is to the I_k .
2. If $P(i, j) = 1$ (i.e., white pixel), then $I_k = \overline{M_n}$. This also implies that there is 50 % probability that white pixel passes as it is to the I_k .

Thus, it makes some kind of confusion about the nature of the original pixel with the compromise in contrast value. Hence, some distortion is introduced in the intermediate matrices. The n shares for n participant from the intermediate matrices are generated using a cover image C according to the following rule:

for $k = 1-n$

$$S_k = C \oplus I_k$$

endfor

Algorithm 2: Algorithm for (2, n) Extended VSS

// Distribution Phase

Pre-condition: Halftoned Image P ; A Gray scale Cover Image: C

Post-condition: Two shares p_1 and p_2 .

- (1) Generate n random binary matrices $M_1, M_2, \dots, M_n$
- (2) //Generate intermediate matrices according to the following rule
 - for $i = 1$ to N
 - for $j = 1$ to N
 - if $P(i, j) == 0$ // black pixel
 - for $k = 1$ to n
 - $I_k(i, j) = M_n(i, j) \oplus M_k(i, j)$
 - endfor
 - else
 - for $k = 1$ to n
 - $I_k(i, j) = \overline{M_n(i, j)}$
 - end for
 - endif
 - end for
- (3) // Generate shares according to the rule
 - for $k = 1$ to n
 - $S_k = C \oplus I_k$
 - end for
- (4) //Distribute cover images and shares

A single share is consist of pair (C, S_i) .

The shares $(C, S_i) | 1 \leq i \leq n$ is distributed to n participants.

// Reconstruction Phase

INPUT: Any two shares (C, S_i) and (C, S_j) , where $i \neq j$.

OUTPUT: Secret image R .

- (1) $I_i = C \oplus S_i$
- (2) $I_j = C \oplus S_j$
- (3) $R = I_i \oplus I_j$

Fig. 3 Algorithm for share generation and revealing phase

Since cover image is a grayscale image, the XORing of C with corresponding I_k has very little effect as XORing only effects the LSB of cover image. This operation generates a single share. The process is repeated n times to generate n shares. Finally, a single share with doublet (C, S_i) is distributed to participant p_i . During revealing phase, any two participants p_i and p_j perform the operation on their shares $(C \oplus S_i)$ and $(C \oplus S_j)$, respectively, to yield intermediate matrices where $1 \leq i, j \leq n$ and $i \neq j$. Finally, any two distinct participants perform the XNOR operation to reveal the secret. Algorithm 2 for generating and revealing the secret image is shown in Fig. 3.

4 Experimental Results and Analysis

This section illustrates the results of an experiment conducted on a grayscale image of Leena (512×512) with a cover image of woman.tif (512×512) shown in Fig. 4a, b. The first step is to transform the grayscale image into an approximate binary image; we used the algorithm discussed in Sect. 3.1. The converted image of size 512×512 pixels is shown in Fig. 4c. Algorithm 2 applied on halftone image

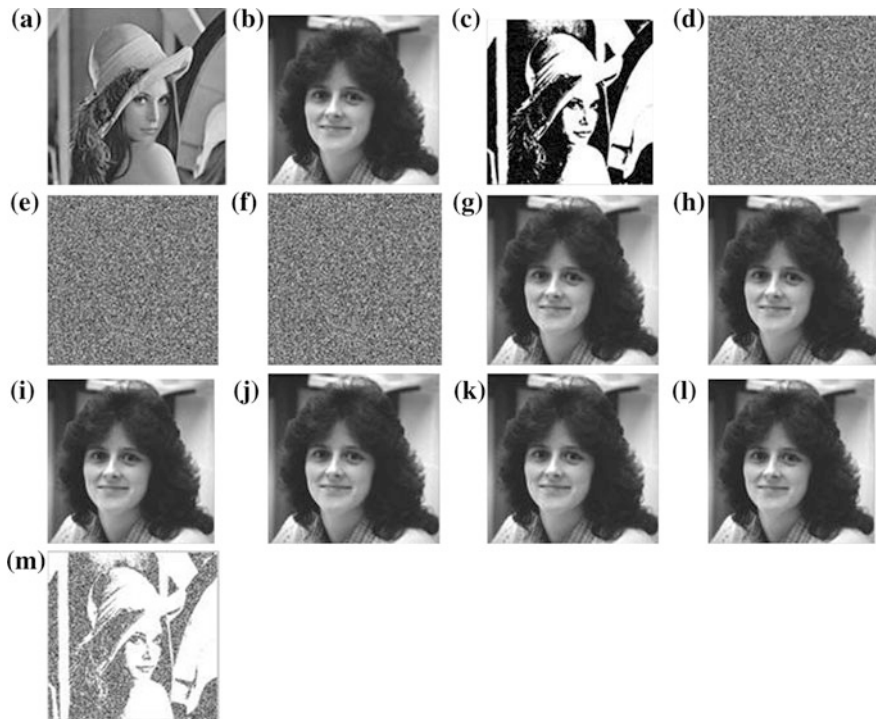


Fig. 4 Result of halftoning and share generation. **a** Secret image. **b** Cover image: C . **c** Halftone image. **d** Intermediate image: $I1$. **e** Intermediate image: $I2$. **f** Intermediate image: $I3$. **g** Share 1: $S1$. **h** Share 2: $S2$. **i** Share 3: $S3$. **j** Participant 1: $(C, S1)$, **k** Participant 2: $(C, S2)$. **l** Participant 3: $(C, S3)$. **m** Revealed image

produces images shown in part d–l of Fig. 4. The intermediate images shown in part d–f are temporary images. These images with the cover image produces shares shown in part g–i. The participants receive a doublet consisting of a cover image and a single share S_i . Once a share S_i is distributed, it cannot be distributed again. At reconstruction, the secret image is obtained by performing XNOR operation. The reconstructed image is shown in Fig. 4m. The experimental results show that our proposed scheme has the following observations:

1. Since encoding and decoding is done on pixels without expanding them, there is no pixel expansion ($m = 1$).
2. The quality of the shadow images are meaningful. This feature is introduced with the help of cover image.
3. The halftoning method explained in Sect. 3 is used to transform the gray level image into a monochrome image. During the probabilistic share generation, the probabilities of sending black-and-white pixels are nearly 50 %. This introduces noise in the secret image which is hidden by some cover image. Thus, when the secret is revealed it mainly suffers due to low PSNR and MSE. The PSNR and MSE values between original secret image and revealed secret image for Fig. 4 are 5.7081 and 17,469. The resulting high value of mean square error is due to both halftoning and share generation phase.
4. As security issue is concerned, it is hard to visualize any difference between individual shares of a participant doublet. However, C and S_i look the same and even though participant p_i makes XOR operation between its components C and S_i , he will never get any information behind the image. To gain complete knowledge of secret image, participation of another participant is mandatory. This makes the scheme more robust and ensures meaningful image.

5 Conclusion

This paper explains the scheme that hides the randomness appeared in the shares by introducing some meaningful information. Visualizing the meaningful information still keeps the actual secret data safe. Such a scheme is known as extended visual cryptography (EVC). This paper introduces an extended cryptography scheme with cover images. The preprocessing of the image is done with the help of the proposed halftoning scheme. The obtained halftoned image is transformed into multiple shares that are distributed to the participants. These shares are finally covered with some cover images to obtain meaningful shares. The experimental results and analysis shows that the proposed scheme has satisfactory results in terms of pixel expansion and security. However, there is scope to develop some improved algorithms that result in low mean square error that arises due to both halftoning and share generation phase.

References

1. Shamir, A.: How to share a secret. *Commun. ACM* 22(11) (nov 1979) 612–613.
2. Naor, M., Shamir, A.: *Visual cryptography*, Springer-Verlag (1995) 1–12.
3. Naor, Shamir: Visual cryptographyii: Improving the contrast via the cover base. In: *Security Protocols Workshop*. Volume 1189. (1996) 197–202.
4. G. Ateniese, C. Blundo, A.D.S., Stinson, D.: Extended capabilities for visual cryptography. *Theoretical Computer Science* 250 (2001) 143–161.
5. Rossand, A., Othman, A.A.: Visual cryptography for biometric privacy. *IEEE Transactions on Information Forensic and security* 6(1) (2011) 70–81.
6. N. Askari, C.M., Heys, H.: A novel visual secret sharing schemewithout image size expansion. In: *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. (2012) 1–4.
7. N. Askari, H.H., Moloney, C.: An extended visual cryptography scheme without pixel expansion for halftone images. *IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)* 6(1) (2013) 33–38.
8. Wu, C., Chen, L.: *A Study on Visual Cryptography*. PhD thesis, Institute of Computer and Information Science, National Chiao Tung University, Taiwan (1998).
9. Sharma, S., Priyadarshni, P.: Analysis and design of secure and contrast enhanced secret sharing scheme using progressive visual cryptography. *International Journal of Science and Research (IJSR)* 3(8) (August 2014) 1181–1186.
10. C. Priyanka, T.V.R., Somashekar, T.: Analysis of secret sharing and review on extended visual cryptography scheme. *academia.edu* 1(10) (2012) 43–51.
11. D. S. Tsai, T.C., Horng, G.: On generating meaningful shares in visual secret sharing scheme. *Imag. Sci. J.* 56 (2008) 49–55.
12. Wang, Z., Arce, G.R., Di Crescenzo, G.: Halftone visual cryptography via error diffusion. *Trans. Info. For. Sec.* 4(3) (September 2009) 383–396.
13. Floyd, R.W., Steinberg, L.: An adaptive algorithm for spatial gray scale. In: *Society for Information Display*. Volume 17. (1976) 75–77.
14. Weir, J., Yan, W.: A comprehensive study of visual cryptography. *T. Data Hiding and Multimedia Security* 5 (2010) 70–105.
15. Nakajima, M., Yamaguchi, Y.: Extended visual cryptography for natural images. In: *Proceedings of WSCG*. (2002) 303–310.

Resonant Cavity-Based Optical Wavelength Demultiplexer Using SOI Photonic Crystals

Chandraprabha Charan, Vijay Laxmi Kalyani
and Shivam Upadhyay

Abstract The performance of a demultiplexer is measured in terms of structure size, quality factor, transmission efficiency, and cross talk level. In the proposed paper, we present a novel structure for separating 1.31 and 1.55 μm wavelength corresponding to original (o) band and conventional (c) band, respectively. The structure utilizes a simple resonant cavity with SOI-based photonic crystal structure. The proposed structure is made of a hexagonal lattice of air holes in silicon slab with the refractive index of 3.47. The numerical results show that proposed structure can play an important role in fiber access networks. The footprint of proposed structure is about $35.25 \mu\text{m}^2$ ($7.5 \mu\text{m} \times 4.7 \mu\text{m}$) that make it suitable for photonic integrated circuits. The mean transmission efficiency and cross talk are about 90 % and -20.34 db . The quality factor measured for 1.31 μm and 1.55 μm are 963 and 1291, respectively.

Keywords Photonic crystals (Phcs) • Resonant cavity • Photonic band gap (PBG) • Finite difference time domain (FDTD) method • Silicon on insulator (SOI) • Plane wave expansion (PWE) method

1 Introduction

Since the discovery of photonic crystals (Phcs) in 1987 [1], the optical devices based on Phcs have been receiving greater attention due to their ultracompact structure, high capacity, high performance, high speed, and long life which make them suitable for ultrasmall integration purpose. Phcs have the ability to confine the

Chandraprabha Charan (✉) • V.L. Kalyani • Shivam Upadhyay
Government Mahila Engineering College, Ajmer, India
e-mail: charanchandraprabha@gmail.com

V.L. Kalyani
e-mail: kalyanivijaylaxmi@gmail.com

Shivam Upadhyay
e-mail: shivamupadhyay920@gmail.com

light inside the structure. Phcs also exhibit photonic band gap (PBG) by which Phcs can prohibit the propagation of electromagnetic wave in certain range of frequency [2–4]. Nowadays research increases the attention to develop Phcs-based devices like multiplexers/demultiplexers, add-drop optical filters (ADF), polarization beam splitters, optical switches and channel-drop optical filters, and so on [5, 6]. Phcs have the ability to select different wavelengths by introducing various defects in structure such as heterostructure with ring resonator, resonant cavity, superprism phenomena in filter structure, radius defects, etc. [7].

Recently Phcs-based demultiplexers play an important role in wavelength division multiplexed (WDM) system and fiber-to-the-home (FTTP)-based systems. Phcs-based wavelength demultiplexers have been proposed in several papers: Parvez et al. proposed a wavelength demultiplexer for 1.31 and 1.55 μm wavelengths. The efficiency of transmission is only 20 % and 70 % for 1.31 μm and 1.55 μm , respectively. Also the wavelength 1.31 μm is associated with high cross talk [8]. A hybrid photonic crystal-based demultiplexer based on coupled line defect channels has been proposed by Yusoff. The device has a power efficiency of about 88 % and also the extinction ratios obtained for 1.31 and 1.55 μm are -25.8 db and -22.9 db [9].

In the proposed structure, silicon on insulator (SOI) is used because it provides several advantages like: we confine electromagnetic (EM) wave in SOI horizontal plan and guide it within this plan using photonic crystal structure, respective indexes of SiO_2 , and silicon allow planar EM wave confinement, SiO_2 can act as an effective barrier against diffusion of carriers which are photo generated or injected in the silicon material [6, 10]. According to fabrication point of view, the proposed structure uses Phcs structure with air holes etched in silicon slab because light confinement is better in such structures as compared to silicon rods in air background structures. In this paper, an ultracompact structure with air holes in silicon slab is used. The structure uses 2D photonic crystals (Phcs-based hexagonal structure with $\text{RI} = 3.47$). A simple resonant cavity is utilized for separating the two optical window wavelengths also SiO_2 as a cladding material is used for insulation purpose. The quality factor of proposed structure for 1.31 μm and 1.55 μm is 963 and 1291, respectively. The efficiency of transmission for 1.31 μm and 1.55 μm wavelengths is 87.15 % and 90.52 %, respectively, also cross talk i.e., one of the critical factor in designing a demultiplexer is between -17.87 and -22.81 db.

2 Structure Design and Analysis

The first purpose in designing a demultiplexer is that it should be simple in fabrication so that they should not have any complexity in design and fabrication. Also for integration purpose the dimension of structure should be compact. After that the cross talk level must be low so that the wavelength can be separated with high

accuracy and minimum cross talk level. The quality factor of structure and transmission efficiency also determines the resolution power and accuracy of structure. In the proposed structure, two resonant cavities are introduced for separating 1.31 and 1.55 μm wavelengths. The resonant cavity is created by changing the radius of certain air holes in structure. The resonant cavity couples a particular wavelength from input waveguide to output waveguide.

The structure is composed of two-layer silicon ($\text{RI} = 3.47$) material as a substrate and SiO_2 for cladding. Two dimensional (2D) Phcs with hexagonal lattice structure are chosen because hexagonal symmetry has smaller angle for bending the electromagnetic wave that result in lower losses and scattering inside the structure. It is found using simulations that a hexagonal lattice with $r/a = 3.173$ yields a wide band gap, where $r = 0.11 \mu\text{m}$ is radius of air holes and $a = 0.349 \mu\text{m}$ is lattice constant of structure.

2.1 Layout of Proposed Structure

Figure 1 shows the layout of the proposed structure. It consists of an input waveguide, two resonant cavities, and two output waveguides. The input waveguide and output waveguide are created by removing air holes in structure. Two resonant cavities are created by changing the radius of three air holes along each output waveguide, where each resonant cavity couples a desired resonant wavelength from input waveguide to output waveguide. The first resonant cavity is created along a straight path for demultiplexing 1.31 μm wavelength. The radius of central hole is $0.085 \mu\text{m}$ and the radius of side holes is $0.053 \mu\text{m}$. Again our goal is

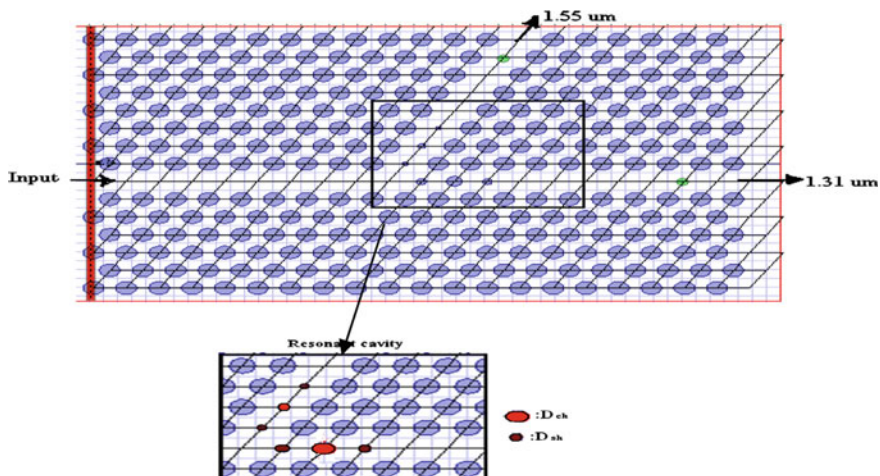


Fig. 1 Layout of proposed structure and resonant cavity

to design a demultiplexer that is capable of separating two wavelengths with high transmission efficiency; to do so the resonant cavities should be different from each other and the second resonant cavity is created along banded path for demultiplexing $1.55 \mu\text{m}$ wavelength. The radius of central hole is $0.04 \mu\text{m}$ and the radius of side holes is $0.027 \mu\text{m}$. In such a way, the desired wavelength is selected with high transmission efficiency and low cross talk level, where each resonant cavity is sensitive to change in radius of air holes along output waveguide.

3 Simulation and Results

After finalizing the structure, Optiwave software tool is used for simulation. The photonic band gap (PBG) is calculated by plane wave expansion (PWE) method and finite difference time domain (FDTD) method is used for numerical computation.

As shown in Fig. 2, the structure has a band gap from $0.21243(a/\lambda)$ to $0.29687(a/\lambda)$ which covers the wavelength range from 1175 to 1642 nm. The Gaussian modulated continuous wave is used for the excitation of the input plane and the perfect matched layer (PML) boundary condition has been used because of its high accuracy and high performance. The structure is composed of 15×20 air holes with structure lying in the x-z plane. The transverse electric (TE) polarization is selected for the propagation of light in z-direction. The structure uses 30,000 time step for simulation. The simulation results for proposed demultiplexer are shown in figures below

The output transmission efficiency obtained for $1.31 \mu\text{m}$ and $1.55 \mu\text{m}$ is 87.15 % and 90.52 %, respectively. The cross talk level for $1.31 \mu\text{m}$ and $1.55 \mu\text{m}$ is -17.87 db and -22.81 db , respectively, which is quite low. In this structure, each resonant cavity is tuned in such a way that minimum cross talk occurs with high transmission efficiency (Fig. 3).

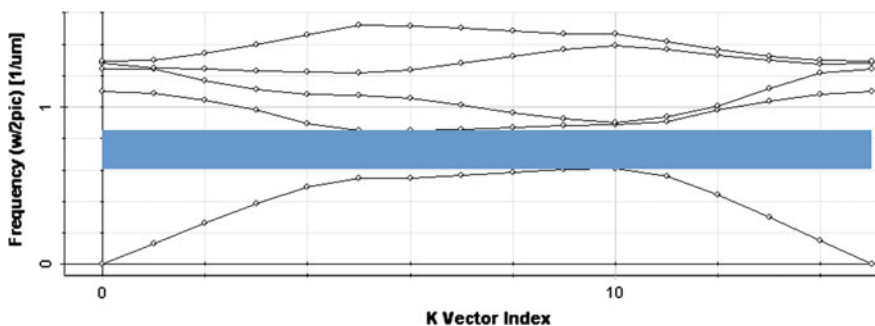


Fig. 2 The band gap for proposed demultiplexer structure

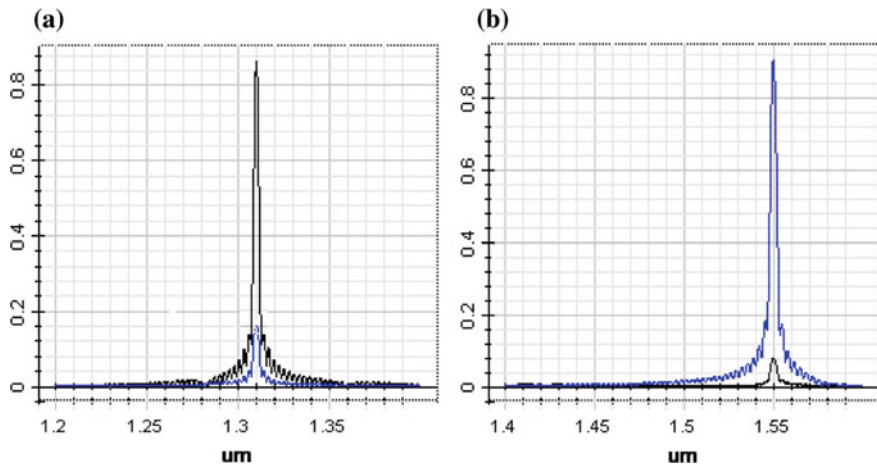


Fig. 3 Transmission power efficiency for **a** 1.31 μm and **b** 1.55 μm

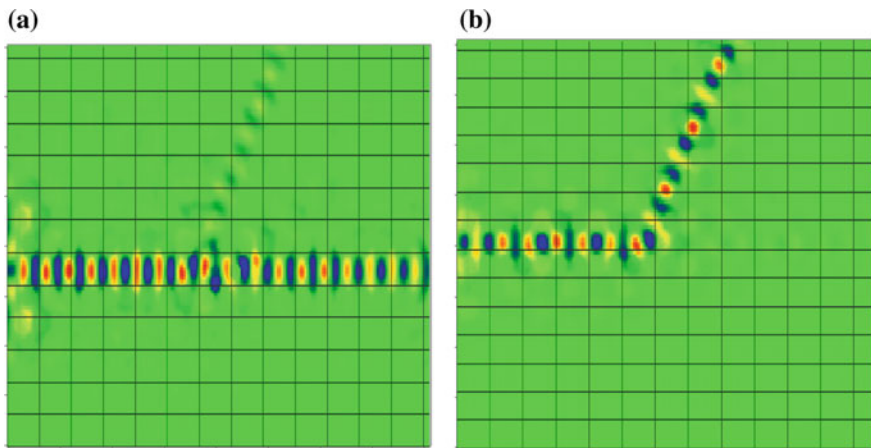


Fig. 4 Steady state field distribution for **a** 1.31 μm and **b** 1.55 μm

Another important parameter that determines the resolution of wavelength selection is quality factor. The quality factor is defined as the ratio of resonant wavelength (λ) to the full width at half power ($\Delta\lambda$) i.e.,

The quality factor for proposed structure is 963 and 1291 for 1.31 μm and 1.55 μm, respectively. Figure 4 shows the FDTD simulated results of the steady state electric field distribution for 1.31 and 1.55 μm (Table 1).

Table 1 Output simulation results for proposed wavelength demultiplexer

Output waveguide	λ (μm)	$\Delta\lambda$ (μm)	Transmission efficiency (%)	Quality factor	Cross talk (dB)
1	1.31	1.36×10^3	87.15	963	-17.87
2	1.55	1.2×10^3	90.52	1291	-22.81

4 Conclusion

In this work, we have demonstrated 2D photonic crystals (Phcs)-based demultiplexer for separating 1.31 and 1.55 μm wavelengths. The device uses a resonant cavity with SOI technology in structure. The total size of structure is $35.25 \mu\text{m}^2$ ($7.5 \mu\text{m} \times 4.7 \mu\text{m}$) i.e., smaller than conventional demultiplexer. The simple and ultracompact structure makes it suitable for fabrication purpose. Again the proposed structure consists of features like high quality factor, high transmission efficiency, and quite low cross talk that make it suitable candidate for FTTP- and WDM-based systems.

Acknowledgment The author would like to thank Mrs. Vijay Laxmi Kalyani Assistant professor in electronics and communication department of Govt. Mahila Engineering College Ajmer for her helpful contribution and guidance and also thanks to references for their literature support

References

1. S. John, "Strong localization of photons in certain disordered dielectric super lattices," Phys. Rev. Lett., vol. 58, pp. 2486–2489, 1987.
2. J. D. Joannopoulos and et al. "Photonics Crystal: Molding the Flow of Light" Princeton University Press, Princeton, NJ, USA, 1995.
3. M. Djavid, M.S. Abrishamian, Multi – channel drop filter using photonic crystals ring resonators, OPTIK - INT. J. LIGHT ELECTRON OPT. 123 (2012) 167–170.
4. S. Robinson, R. Nakkeeran "PCRR based add-drop filter for ITU-T G.694.2 CWDM system" OPTIK - INT. J. LIGHT ELECTRON OPT. In Press (2012).
5. Mohammad Ali mandouri-birjandi and Alireza Tavousi "Design and Simulation of an Optical Channel Drop Filter Based on Two Dimensional Photonic Crystal Single Ring Race Track Resonator" in International Journal of Natural and Engineering Sciences 7 (1): 14–18, 2013 ISSN: 1307-1149.
6. Fan S H, Johnson S G, Joannopoulos J D, Manolatu C and Haus H A 2001 "Waveguide branches in photonic crystal" J. Opt. Soc. Am. B 18 162–5.
7. Ali Rostami and et al. "A novel proposal for DWDM demultiplexer design using resonant cavity in photonic crystals structure" Proc. of SPIE-OSA-IEEE Asia Communications and Photonics, SPIE Vol. 7630, 763013, 2009.
8. Md. Masud Parvez Amoh et al. "Design and Simulation of an Optical Wavelength Division Demultiplexer Based on the Photonic Crystal Architecture" IEEE/OSA/IAPR International Conference on Infonnatics, Electronics & Vision 978-1-4673-1154-0112, 20 12.
9. Mohd hanapian m. yusoff and et al. "Hybrid photonic crystal 1.31/1.55 μm wavelength division multiplexer based on coupled line defect channels" optics communication 280 (2011) 1223–1227.
10. T. Charvolin et al. "Realization of two-dimensional optical devices using PBG structures on silicon-on-insulator" Microelectronic Engineering 61–62 (2002) 545–548.

A Frequency Reconfigurable Antenna with Six Switchable Modes for Wireless Application

Rachana Yadav, Sandeep Yadav and Sunita

Abstract In this paper, a frequency switchable microstrip patch antenna with defected ground structure is presented. Frequency characteristics of this antenna can be switched between different frequency bands. Reconfigurability is achieved using different slots on ground structure and three PIN diodes loaded onto these slots as switches. By changing the PIN diode states to either ON or OFF, any particular slot on ground structure will be activated, hence making the antenna to be operable in six different modes which serves different frequency bands to be used for different wireless applications. Return loss, VSWR and gain are analyzed for different modes of operation. To design and simulate the proposed antenna CST microwave studio is used.

Keywords PIN diode · Reconfigurable antenna · Switching · Wireless applications

1 Introduction

In today's communication system, reconfigurable antenna's multifunctional capability plays an advantageous role. Based on reconfigurability of antenna characteristics it is classified into four basic categories—frequency reconfigurable antenna [1], polarization reconfigurable antenna [2], radiation pattern reconfigurable antenna [3], and hybrid (which is the combination of the any of the three other

Rachana Yadav (✉) · Sandeep Yadav · Sunita
Department of Electronics & Communication Engineering,
Government Women Engineering College, Ajmer, Rajasthan, India
e-mail: rachanayadav2112@gmail.com

Sandeep Yadav
e-mail: Sandeep.y9@gmail.com

Sunita
e-mail: olasunita30@gmail.com

categories) [4]. Operating state of antenna can be realized using mechanical or physical alteration, reconfigurable biasing network or using different types of switches (pin diodes, varactor diodes, MEMS etc.) which alter the surface current distribution and change the antenna characteristic [5, 6].

In last few years, many techniques have been used for frequency reconfigurable antenna. PIN diode as a switch is used more frequently because of easy assembly and low cost. MEMS also have been used more often as they have low insertion losses and low power consumption. In wireless communication system, some specific frequency bands are used only for some specific purposes. A frequency reconfigurable antenna with switchable bands provides only one antenna to operate at different wireless standards. In [7], a frequency reconfigurable antenna with narrowband and dual band characteristics is presented. This antenna operates in WLAN (2.4–2.48 GHz) and WiMAX (2.5–2.69 GHz), while the dual-band covers the PCS (1.85–1.99 GHz) and WiMAX (3.4–3.69 GHz). GaAs field effect transistor is used as a switch to achieve reconfigurability. In [8], a reconfigurable antenna with T-slot in patch and E-slot in ground plane is presented. This T-slot divides antenna into three parts. Two PIN diodes are used for reconfigurability antenna works in triple bands (3.9, 8.9 and 11.2 GHz) in one mode and operates in three other bands (4.1, 8.4 and 11.3 GHz) in another mode. In [9], a frequency and pattern reconfigurable antenna is presented. This antenna has three modes of operation, an omnidirectional pattern mode at the lower frequency band of 2.21–2.79 GHz, a unidirectional pattern mode at the higher frequency band of 5.27–5.56 GHz, and both of them working simultaneously.

In this paper, a microstrip patch antenna with defected ground structure is introduced. Antenna is capable to operate in six different modes. To achieve reconfigurability, three PIN diodes are used as switches. These diodes are mounted on the antenna between different slots. By changing the ON/OFF state of diode, different combination of slots are activated at different mode which alter the current distribution of antenna and make it to resonate at different frequencies.

2 Antenna Configuration and Design

Front view of proposed antenna is shown in Fig. 1. This is a simple microstrip patch antenna with operating frequency 5.3 GHz. Firstly, antenna is designed using transmission model approach of microstrip antenna. After that, slots have been cut in ground structure of different size and shape.

Antenna has a radiating patch with patch length L and width W and FR-4 substrate of dielectric constant 4.4. Length and width of ground and substrate are same for the basic antenna design. Substrate length and height are L_s and W_s respectively. Other parameters of antenna are shown in Table 1.

Simulated return loss of base antenna is shown in Fig. 2. A frequency band around 5.31 GHz is achieved.

Fig. 1 Antenna structure
(Front view)

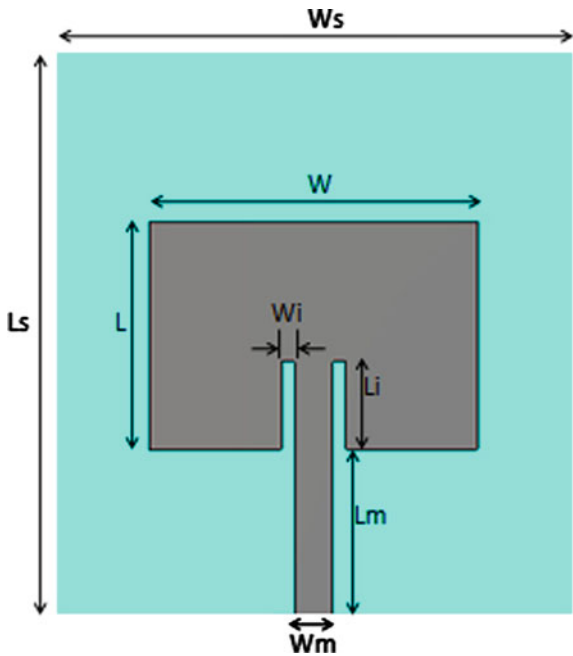


Table 1 Antenna dimensions
of proposed antenna

Parameter	Values (mm)
Width of patch (W)	17.4
Length of patch (L)	12.48
Width of substrate (W_s)	27.48
Length of substrate (L_s)	30.8
Dielectric constant (ϵ_r)	4.3
Height of the substrate (h)	1.67
Height of the patch and ground	0.05
Microstrip feed length (L_m)	9
Microstrip feed width (W_m)	2
Inset width (W_i)	0.7
Inset length (L_i)	4.8

3 Frequency Reconfigurable Antenna Configuration and Design

Now, to get switchable multiple frequencies, slots are cut in ground plane. From the Fig. 3 it can be seen that slots are not of symmetrical shape. Change in slots geometry has been made to make antenna resonate at some specific frequencies.

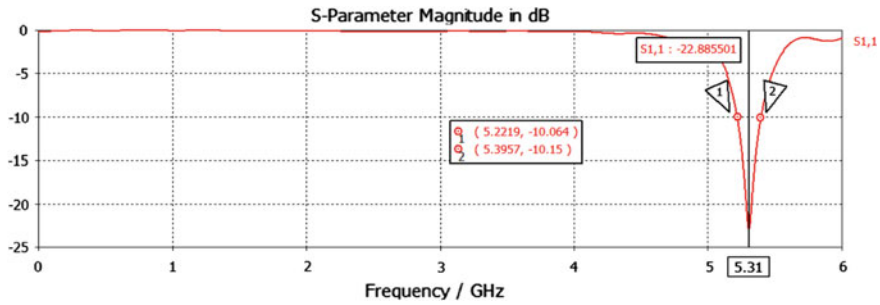
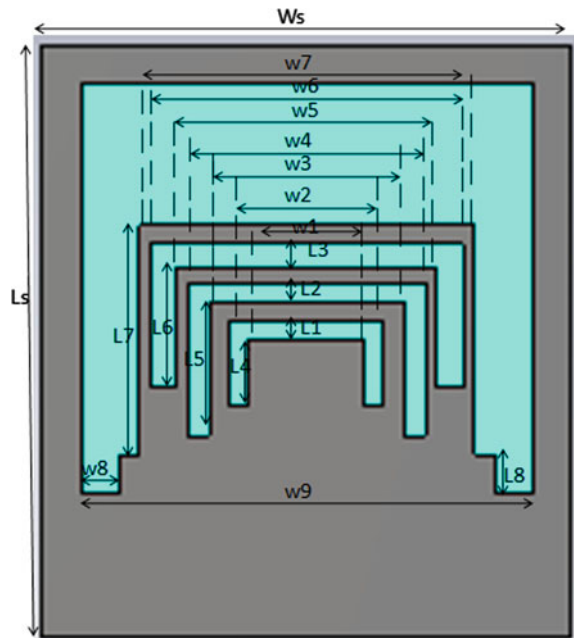


Fig. 2 Return loss of base microstrip patch antenna

Fig. 3 Proposed frequency reconfigurable antenna without switch (*Back-view*)



Three switches are loaded in slots. Here, HPND-4005 beam lead PIN diodes are used as a switch. In ON state, diode is modeled by a 1.5Ω register and in OFF mode modeled by a 0.017 pF capacitor (Fig. 4 and Table 2).

3.1 Simulated Results

Based on the ON/OFF state of the PIN diode and using different combination of these switches' state proposed antenna will work in six modes. At particular

Fig. 4 Proposed frequency reconfigurable antenna with switches (*Back-view*)

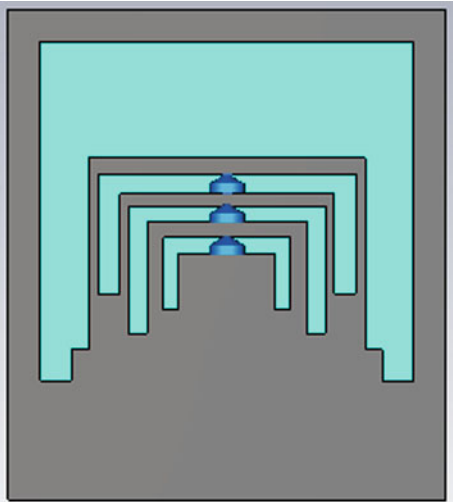


Table 2 Dimensions of slots of ground plane

W1	W2	W3	W4	W5	W6	W7	W8	W9	Ws
6	8	10	12.4	13.4	16.2	17.4	1.96	23.42	27.42
L1	L2	L3	L4	L5	L6	L7	L8	Ls	
1	0.98	1.25	3.5	7	6.1	11.98	2	30.8	

instance, when ON/OFF state of three diodes take place, then according to the activation of the one or more slots, surface current distribution will change which then affects the resonant frequency. This is the basic phenomenon of switching here (Figs. 5, 6, 7, 8, 9, 10, 11 and Tables 3 and 4).

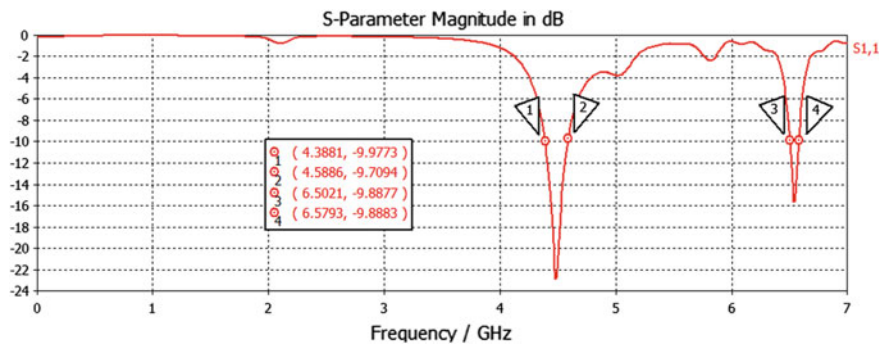


Fig. 5 Return loss plot for mode I

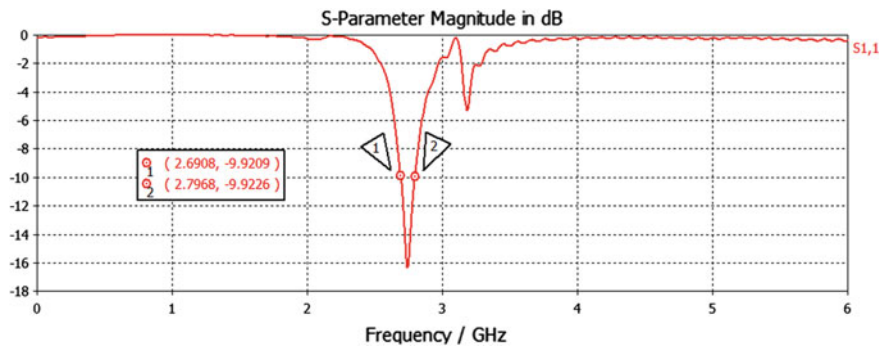


Fig. 6 Return loss plot for mode II

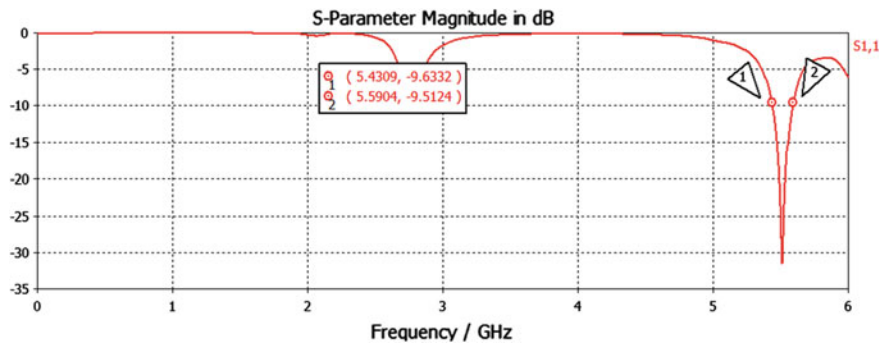


Fig. 7 Return loss plot for mode III

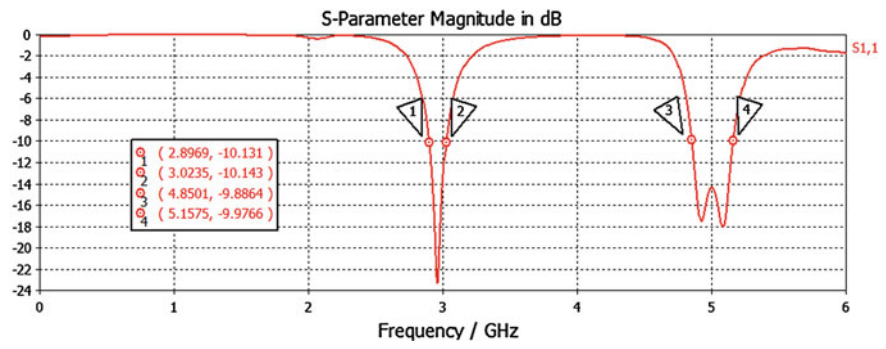


Fig. 8 Return loss plot for mode IV

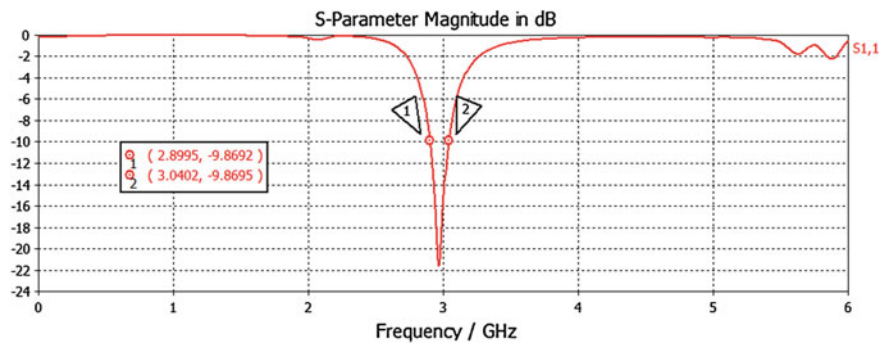


Fig. 9 Return loss plot for mode V

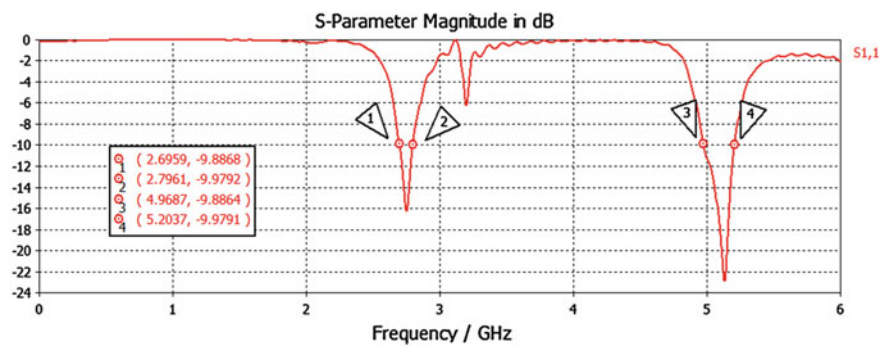


Fig. 10 Return loss plot for mode VI

4 Results and Discussion

To design and simulate the proposed frequency reconfigurable antenna, CST microwave studio is used. Reconfigurable frequency bands are achieved using ON/OFF states of the PIN diodes. Seven different frequency bands are obtained with center frequency 4.4 GHz (Rx frequency for INSAT), 6.5 GHz (Tx frequency for extended c-band), 2.7 GHz (WISP/NLOS/802.16), 5.5 GHz (UI-wireless), 2.9 GHz, 5 GHz (Wi-fi and WLAN), and 5.12 GHz (UNII-1) 12.3 GHz. Return loss and gain has been analyzed for individual mode and presented systematically. Because of its multifunctional property, antenna is operable for multiple wireless applications.

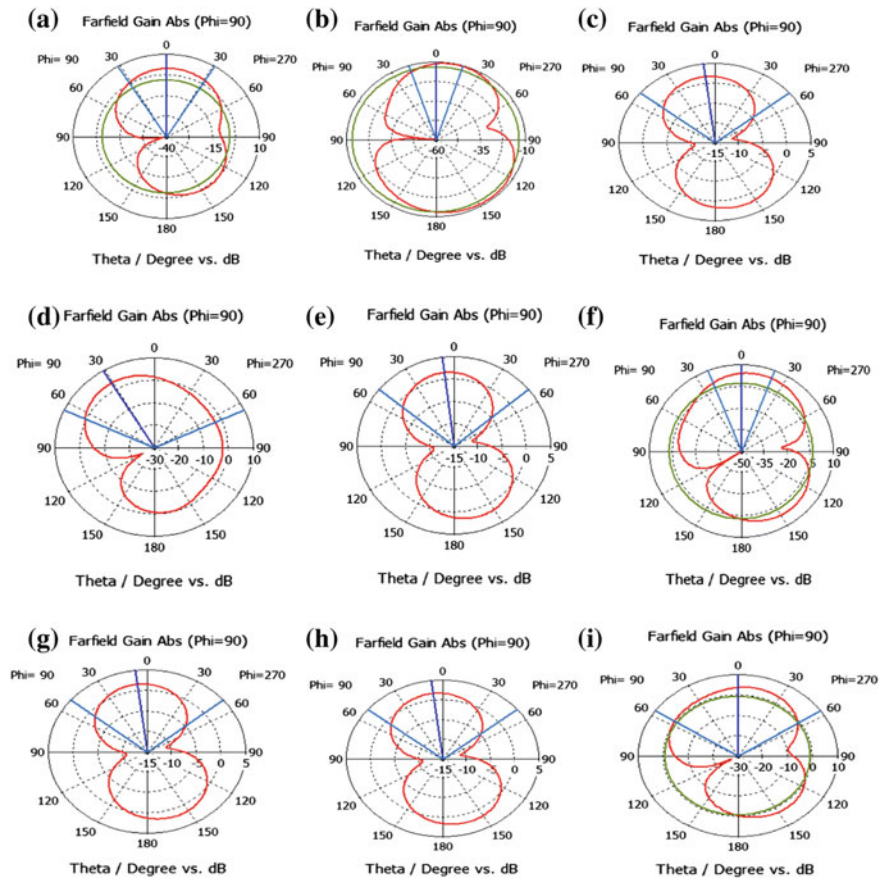


Fig. 11 Gain for six modes—**a, b** for mode I, **c** for mode II, **d** for mode III, **e, f** for mode IV, **g** for mode V, **h, i** for mode VI

Table 3 Summarized results for proposed antenna for OFF state and ON state (in terms of operating frequency/ies)

Mode	Switch state			Center frequency (bands in GHz)
	S1	S2	S3	
I	ON	ON	ON	4.47 GHz (4.38–4.58), 6.53 GHz (6.50–6.57)
II	ON	OFF	OFF	2.73 GHz (2.69–2.79)
III	ON	ON	OFF	5.5 GHz (5.43–5.59)
IV	OFF	OFF	ON	2.95 GHz (2.89–3.02), 5 GHz (4.85–5.15)
V	ON	OFF	ON	2.96 GHz (2.89–3.02)
VI	OFF	OFF	OFF	2.74 GHz (2.69–2.79), 5.12 GHz (4.96–5.20)

Table 4 Summarized return loss and gain results

Mode	Return loss	Gain (dB)
I	−23 dB at 4.47 GHz, −15 dB at 6.53 GHz	1.1 and 3.5
II	−16 dB 2.73 GHz	1.9
III	−32 dB at 5.5 GHz	3.3
IV	−23 dB at 2.95 GHz, −16 dB at 5 GHz	1.6, 4.1
V	−22 dB at 2.96 GHz	1.7
VI	−16 dB at 2.74 GHz, 5.12 GHz at −23 dB	1.9, 3.8

References

1. Tummas, P.; Krachodnok, P.; Wongsan, R., "A frequency reconfigurable antenna design for UWB applications," *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2014 11th International Conference on*, vol., no., pp. 1, 4, 14–17 May 2014.
2. Wang, G.; Bairavasubramanian, R.; Pan, B.; Papapolymerou, J., "Radiofrequency MEMS-enabled polarisation reconfigurable antenna arrays on multilayer liquid crystal polymer," *Microwaves, Antennas & Propagation, IET*, vol. 5, no. 13, pp. 1594, 1599, October 2011.
3. Adhikari, M.; Warnick, Karl F., "Miniature radiation pattern reconfigurable antenna for 2.4 GHz band," *Antennas and Propagation Society International Symposium (APSURSI), 2010 IEEE*, vol., no., pp. 1, 4, 11–17 July 2010.
4. Rodrigo, D.; Cetiner, B.A.; Jofre, L., "Frequency, Radiation Pattern and Polarization Reconfigurable Antenna Using a Parasitic Pixel Layer," *Antennas and Propagation, IEEE Transactions on*, vol. 62, no. 6, pp. 3422, 3427, June 2014.
5. Kiriazi, J.; Ghali, H.; Ragaie, H.; Haddara, H., "Reconfigurable dual-band dipole antenna on silicon using series MEMS switches," *Antennas and Propagation Society International Symposium, 2003. IEEE*, vol. 1, no., pp. 403, 406, 22–27 June 2003.
6. Ramli, N.; Ali, M.T.; Yusof, A.L.; Muhamud-Kayat, S.; Aziz, A.A.A., "PIN diode switches for frequency-reconfigurable stacked patch microstrip array antenna using aperture-coupled technique," *Microwave Conference Proceedings (APMC), 2013 Asia-Pacific*, vol., no., pp. 1052, 1054, 5–8 Nov. 2013.
7. Xiao-lin Yang; Jian-cheng Lin; Gang Chen; Fang-ling Kong, "Frequency Reconfigurable Antenna for Wireless Communications Using GaAs FET Switch," *Antennas and Wireless Propagation Letters, IEEE*, vol. 14, no., pp. 807, 810, Dec. 2015.
8. Kumari, R.; Kumar, M., "Frequency reconfigurable multi-band inverted T-slot antenna for wireless application," *Advances in Computing, Communications and Informatics (ICACCI), 2014 International Conference on*, vol., no., pp. 696, 699, 24–27 Sept. 2014.
9. Peng Kai Li; Zhen Hai Shao; Quan Wang; Yu Jian Cheng, "Frequency- and Pattern-Reconfigurable Antenna for Multistandard Wireless Applications," *Antennas and Wireless Propagation Letters, IEEE*, vol. 14, no., pp. 333, 336, 2015.

Comparative Analysis of Digital Watermarking Techniques

Neha Bansal, Vinay Kumar Deolia, Atul Bansal and Pooja Pathak

Abstract In this paper various techniques used for digital watermarking such as least significant bit (LSB) technique, discrete cosine transform (DCT), discrete wavelet transform (DWT), and back propagation neural network (BPN) algorithm have been compared. These techniques are used to embed and extract a watermark of an image. The performance of these algorithms is evaluated using various parameters such as mean square error, peak signal-to-noise ratio (PSNR), and normalized correlation (NC). Parameters for each technique are compared for various noises like Gaussian noise, Poisson noise, salt-and-pepper noise, and speckle noise. Based on comparison it is suggested that BPN gives better result in terms of PSNR and NC.

Keywords Digital watermarking • Least significant bit (LSB) technique • Discrete fourier transform (DFT) • Discrete cosine transform (DCT) • Discrete wavelet transform (DWT) • Back propagation neural network (BPN) • Counter propagation neural network (CPN) • Normalized cross-correlation (NC) • Peak signal-to-noise ratio (PSNR)

1 Introduction

Digital watermarking is a method to prevent illegal copying of digital content as it can be copied and edited easily. Digital watermarking can be done in various ways. It can be done in spatial domain using least significant bit (LSB) technique. It can also be done in spectral domain using various transforms such as discrete fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet transform (DWT). Another method of digital watermarking is based on neural network.

Neha Bansal (✉) · V.K. Deolia · Atul Bansal
Department of ECE, GLA University, Mathura, India
e-mail: neha.bansal@gla.ac.in

Pooja Pathak
Department of Mathematics, GLA University, Mathura, India

Various types of neural network algorithm like back propagation neural network, counter propagation neural network, etc., can be used for it. This method is highly secure because in this method, watermarked image is not sent so it cannot be harmed.

2 Classification of Digital Watermarking Schemes

Various types of watermarking methods are used for the protection of digital data. Some of which are:

2.1 Spatial Domain Watermarking Technique

In spatial domain, watermarking is done in pixel domain. The pixel domain methods have main strengths that they are theoretically simple and have very less computational complexities. Embedding of the watermark into cover image is based on the operations like shifting or replacing of bits. Most commonly used spatial domain watermarking technique is least significant bit technique. In this technique, pixel values of cover image as well as watermark image are converted into binary form. The bits of watermark image replace the least significant bit of cover image and in this way, watermark can be embedded into cover image. Figure 1 shows the framework of the embedding using LSB technique.

The extraction is also very simple. Watermark data can be extracted by matching the supposed sample with the received data. At the extractor end, a zero matrix equal to the size of watermark is taken for the purpose of extraction. Each element

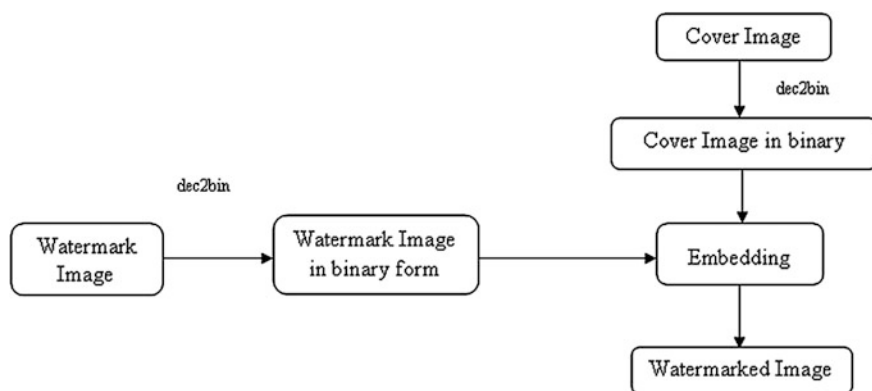


Fig. 1 Embedding of watermark using LSB technique

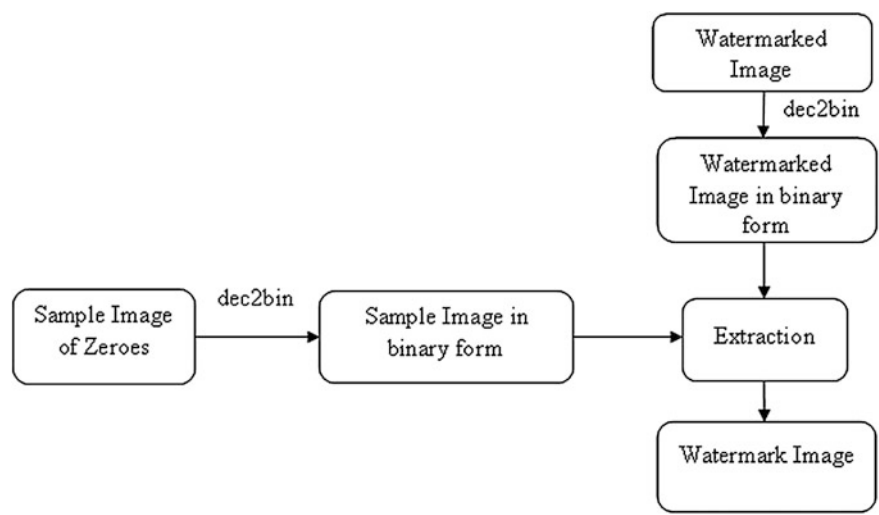


Fig. 2 Extraction of watermark using LSB technique

of zero matrix is converted into binary form as well as watermarked image pixels are also converted into binary form. The least significant bits of watermarked image are replaced by each bit of zero matrix. In this way watermark is retrieved by the extractor. Figure 2 shows the framework of the extraction using LSB technique.

In the proposed method, the cover image is of size $m \times n$ and the watermark image is of size $(m \times n)/8$. The 8th bit of each pixel of cover image is replaced by each bit of the watermark image. The 8th bit of a binary number has least significance so its effect on the cover image is minimum. In this way watermark is embedded and watermarked image is obtained. The performance will be measured using MSE, peak signal-to-noise ratio (PSNR), and normalized correlation (NC). The process is shown in Fig. 3.

2.2 Spectral Domain Watermarking Technique

2.2.1 Watermarking Using DCT

The DCT is a very favored transform function used in digital signal processing. DCT can also be applied in pattern recognition, data compression, and image processing.

Figure 4 shows the framework of the embedding using DCT. Digital watermarking can be done by applying DCT on cover image to get transformed coefficients. If cover image coefficient is represented as C_a , W_i is the corresponding bit of

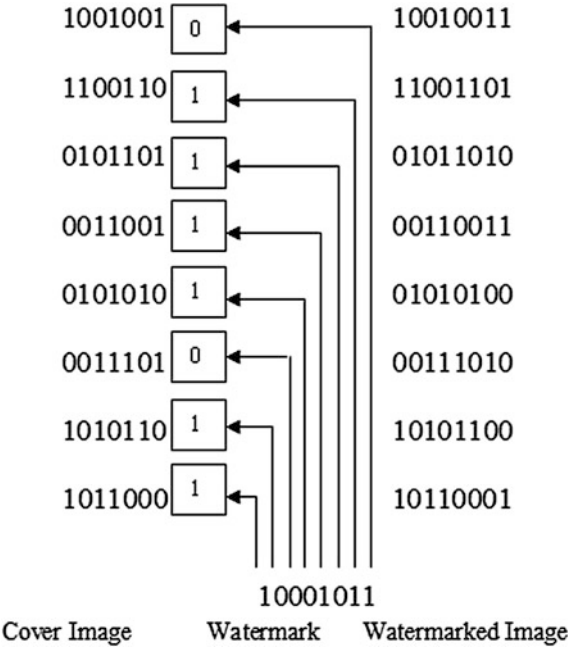


Fig. 3 Process of LSB watermarking using 8th bit

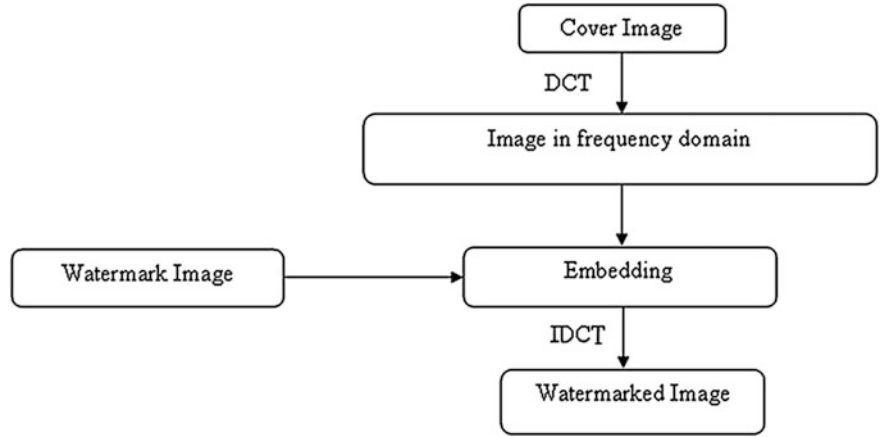


Fig. 4 Embedding of watermark using DCT

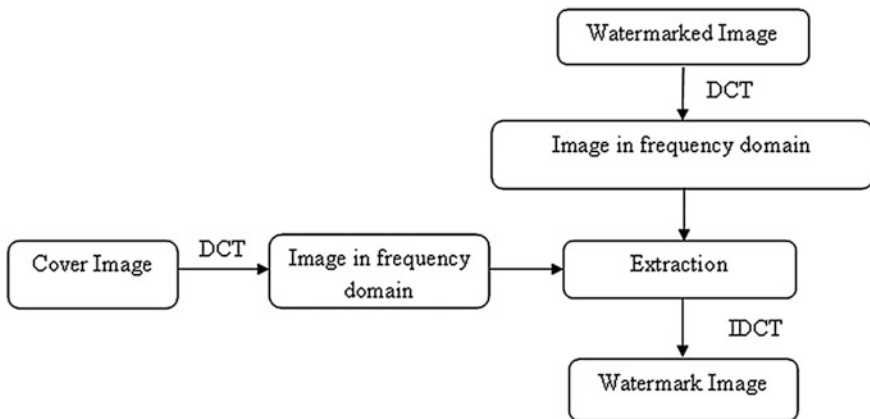


Fig. 5 Extraction of watermark using DCT

the message data, α denotes watermarking strength, and watermarked coefficient is represented as C_{aw} then coefficients are altered depending upon the stream bits of the message using the equation

$$C_{aw} = C_a(1 + \alpha W_i) \quad (1)$$

Figure 5 shows the framework of the extraction using DCT. The extraction can be done in reverse manner. The extracted image can be obtained depending upon the difference between the original DCT coefficients and the watermarked image ones. It can be obtained by the following formula:

$$W_i = \frac{1}{\alpha}(C_{aw} - C_a) \quad (2)$$

2.2.2 Watermarking Using DWT

Wavelet technique is another significant domain for watermarking. When DWT is applied to an image, it decomposes the image into four significant components which are lower resolution (LL), horizontal (HL), vertical (LH), and diagonal (HH) detail components. Figure 6 shows the framework of the watermark embedding using DWT. Watermarking using DWT can be done by applying DWT on cover image to decompose it into four parts. If cover image coefficient is represented as C_a , it is decomposed into four parts, W_i is also decomposed into four parts, α represents watermarking strength and watermarked decomposition is

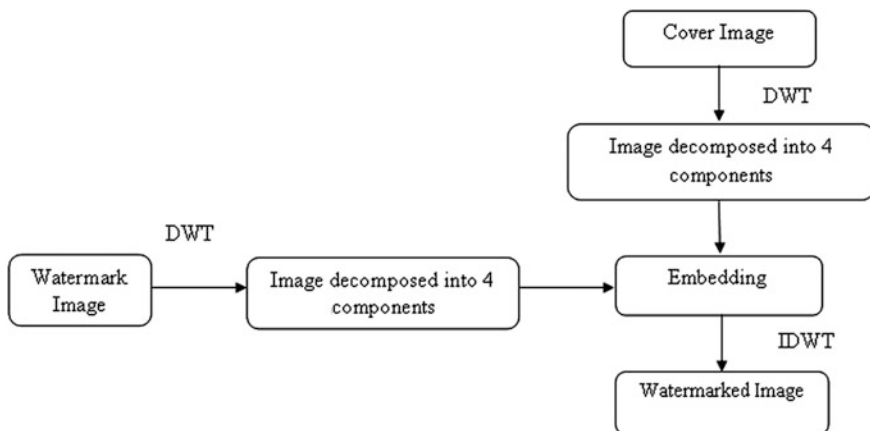


Fig. 6 Embedding of watermark using DWT

represented as C_{aw} then coefficients are altered depending upon the stream bits of the message using equation

$$C_{aw} = C_a(1 + \alpha W_i) \quad (3)$$

Figure 7 shows the framework of the extraction using DWT. The extraction can be done in reverse manner. The extraction can be done by subtracting the original

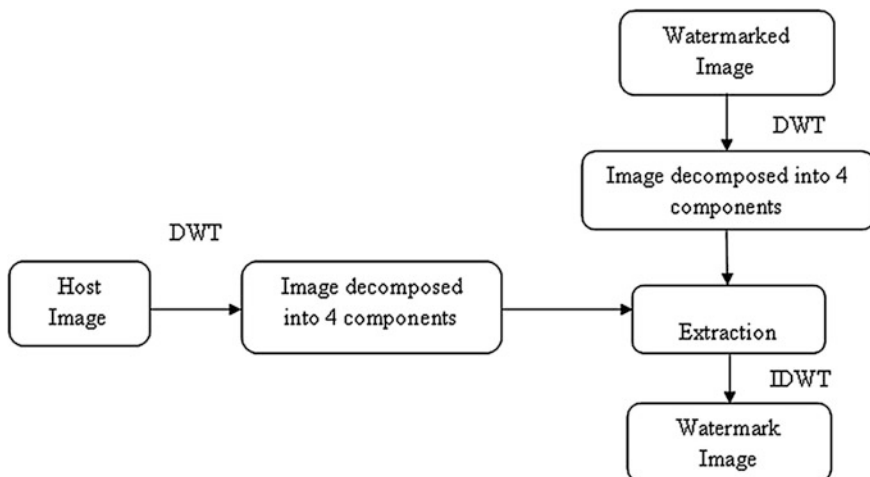


Fig. 7 Extraction of watermark using DWT

DWT coefficients from the watermarked image ones. It can be obtained by the following formula:

$$W_i = \frac{1}{\alpha}(C_{aw} - C_a) \quad (4)$$

2.2.3 Watermarking Using Back Propagation Neural Network

Digital watermarking can be done using back propagation neural network (BPN). BPN can be used to embed the watermark as well as to extract the watermark.

Embedding of watermark using BPN can be done using following steps:

- The cover image and watermark image are divided into small fragments of size 2×1 .
- A BPN is taken with input layer, one hidden layer, and output layer.
- The fragments of cover image are supplied as input to the BPN and the network is trained to generate the fragments of watermark image. Weights are adjusted to produce the desired output for the given input.
- The weights are stored in a file and the cover image with the weight file is sent to the extractor.

The process of watermark embedding is shown in Fig. 8. Extraction of watermark using BPN can be done using following steps:

- At the extractor end, both files are received (weight file and cover image).
- The cover image is divided into small fragments of size 2×1 .
- The weights are extracted from the weight file and BPN is reconstructed.
- With the help of fragments of cover image and trained weights, BPN gives the output same as watermark image.

The process of watermark extraction is shown in Fig. 9.

The performance of this technique is also measured for noised image. Various types of noises are used such as Gaussian noise, Poisson noise, salt-and-pepper noise, and speckle noise.

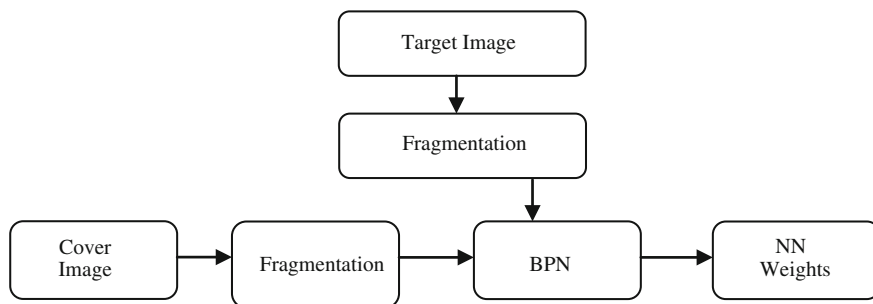


Fig. 8 Watermark embedding using BPN

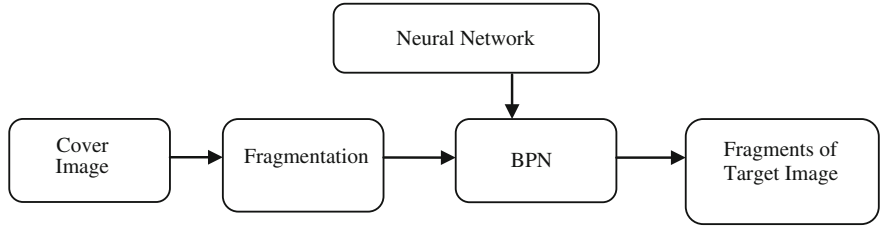


Fig. 9 Watermark extraction using BPN

3 Results

Digital watermarking can be done using various techniques. Watermark is embedded in cover image and the embedded image is sent to the receiver. PSNR and NC give the robustness of the technique. The time consumed by different techniques has been also compared in this work. The results obtained are as follows: (Tables 1, 2 and 3) (Figs. 10, 11 and 12).

Table 1 PSNR values for digital watermarking techniques (dB)

Methods	LSB in 6th bit	LSB in 7th bit	LSB in 8th bit	DCT	DWT	BPN
Without noise	137.75	151.33	165.81	109.2	59.8	129.36
Gaussian noise	46.193	46.321	46.367	66.26	44	115.35
Poisson noise	63.998	63.945	63.919	63.86	54.24	129.07
Salt-and-pepper noise	41.848	42.397	42.352	42.27	40.48	85.474
Speckle noise	47.605	47.629	47.655	47.59	43.85	129.36

Table 2 NC values for digital watermarking techniques

Methods	LSB in 6th bit	LSB in 7th bit	LSB in 8th bit	DCT	DWT	BPNN
Without noise	1	1	1	1	0.994	1
Gaussian noise	0.73	0.688	0.6918	1	0.447	1
Poisson noise	0.6388	0.7005	0.7282	1	0.834	1
Salt-and-pepper noise	0.978	0.9955	0.995	1	0.968	0.999
Speckle noise	0.6522	0.7274	0.733	1	0.741	1

Table 3 Time consumed in various digital watermarking techniques

Methods	LSB in 6th bit	LSB in 7th bit	LSB in 8th bit	DCT	DWT	BPN
Without noise	0.4695	0.4765	0.2822	1.169	1.762	947.57
Gaussian noise	0.6002	0.6115	0.5692	1.311	1.19	779.99
Poisson noise	0.6131	0.6026	0.5743	1.056	1.2	1811.5
Salt-and-pepper noise	0.5907	0.6078	0.5943	1.078	1.203	1999.8
Speckle noise	0.6078	0.5927	0.6121	1.196	1.271	750.24

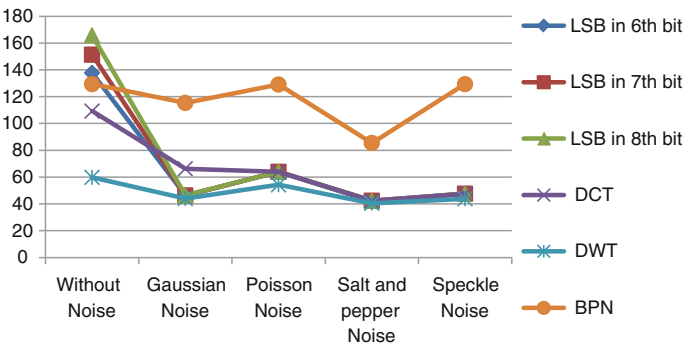


Fig. 10 Graphical representation of PSNR for the proposed technique

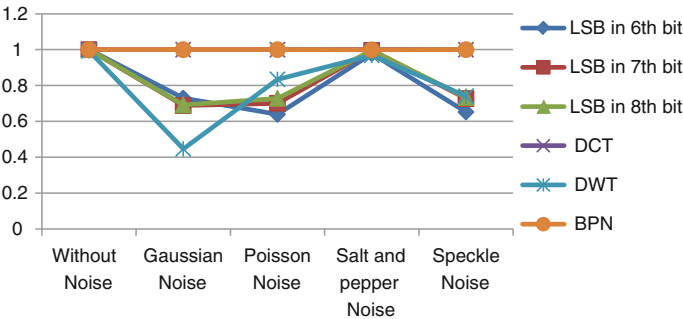


Fig. 11 Graphical representation of NC for the proposed technique

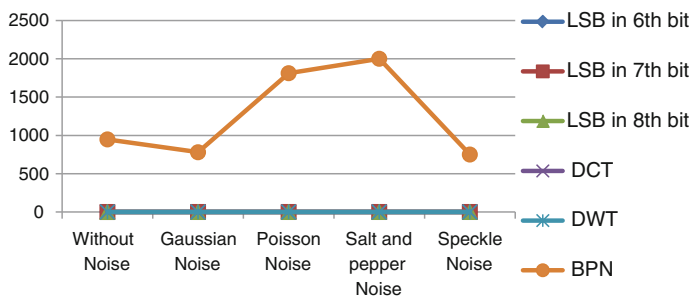


Fig. 12 Graphical representation of time consumed for the proposed technique

4 Conclusion

In this work LSB, DCT, DWT, and BPN are used to embed the watermark with cover image which is being sent to the extractor. The performance has been evaluated using PSNR and NC. On the basis of above results, it is clear that spatial domain is the easiest method but it is less secure. Watermarking using DCT and DWT is more robust. The results of watermarking using BPN are best and it is robust as well as secure technique. But the time consumed in BPN technique is higher than in other techniques.

5 Future Work

This work can be further developed using high security algorithms for embedding and extraction of watermark using full counter propagation neural network, etc.

References

1. C.Y. Chang and S.J. Su, "The Application of a Full Counterpropagation Neural Network to Image Watermarking", in Proc. of IEEE Networking, Sensing and Control, pp. 993–998, 2005.
2. Wai C. Chu, "DCT-Based Image Watermarking Using Subsampling", IEEE Trans. on multimedia, vol. 5, no. 1, pp. 34–38, March 2003.
3. Puneet Kr Sharma and Rajni, "Information security through Image Watermarking using Least Significant Bit Algorithm," Computer Science & Information Technology, vol. 2, no. 2, May 2012.
4. Gaurav Bhatnagar and Balasubramanian Raman, "A new robust reference watermarking scheme based on DWT-SVD", Elsevier Computer Standards & Interfaces, vol. 31, no. 5, September 2009, pp. 1002–1013.

5. Neha Bansal and Pooja Pathak, "A Review of Applications of Neural Network In Watermarking" In the Proc. of 7th National Conference on Advancement of Technologies-Information Systems & Computer Networks, pp. 9–12.
6. Neha Bansal, Vinay Deolia, Atul Bansal and Pooja Pathak, "Digital Image Watermarking Using Least Significant Bit Technique in Different Positions", In the Proc. of 6th International Conference on Computational Intelligence and Communication Networks, pp. 813–818, 2014.
7. Neha Bansal, Vinay Deolia, Atul Bansal and Pooja Pathak, "Comparative Analysis of LSB, DCT and DWT for Digital Watermarking", In the Proc. of the 2nd International Conference on "Computing for Sustainable Global Development", pp. 2.1–2.6, 2015.
8. Mauro Barni, Franco Bartolini, Vito Cappellini and Alessandro Piva, "A DCT-domain system for robust image watermarking," Elsevier Signal Processing 66, 1998, pp. 357–372.
9. J.J.K. O'Ruanaidh, W.J. Dowling and F.M. Boland, "Watermarking digital images for copyright protection", IEE Proceedings—Vision, Image and Signal Processing, vol. 143, no. 4, August 1996, pp. 250–256.
10. M.D. Swanson, M. Kobayashi and A.H. Tewfik, "Multimedia data-embedding and watermarking techniques", Proc. of IEEE, vol. 86, no. 6, 1998, pp. 1064–1087.

Design and Analysis of 1×6 Power Splitter Based on the Ring Resonator

Juhi Sharma

Abstract In this paper, the design and performance of two-dimensional (2D) photonic crystal (PhC) T-shaped 1×6 power splitter based on the ring resonator on square lattice are presented. The coupling characteristic between the waveguide and the ring resonator is analyzed theoretically by the coupled mode theory (CMT). The simulation result of the splitting properties of the T-shaped splitter is obtained numerically by the finite difference time domain (FDTD) method. The uniform splitting can be achieved on the both sides of input waveguide due to the symmetry of the structure. The photonic band gap (PBG) is calculated by the plane wave expansion (PWE) method. The number of rods is 27×42 in x - z plane. The device is ultracompact with the overall size around $322 \mu\text{m}^2$. The photonic crystal power splitter based on the ring resonator is designed for photonic integrated circuits application.

Keywords Photonic crystal • Coupled mode theory • FDTD • PWE • PBG • PCRR • DFT • PML • Two-dimensional

1 Introduction

The photonic crystals (PhC) allow the control of photons similar to the semiconductors that allow the control of electrons. The photonic crystals consist of periodic dielectric nanostructures that affect the propagation of electromagnetic waves. Yablonovitch [1] and John [2] proposed the idea that periodic dielectric structures are able to provide photonic band gap (PBG) for distinct regions in the frequency spectrum similar to the electronic band gap in solid state crystal behavior. There are three types of photonic crystal; one-dimensional, two-dimensional, and three-dimensional crystals, which depend upon the variation of dielectric constant in one, two, and three directions, respectively. The photonic band gap (PBG) is the

Juhi Sharma (✉)

Department of ECE, Government Engineering College, Ajmer, India
e-mail: juhis24@gmail.com

region where the propagation of light is completely prohibited in certain frequency ranges. Photonic band gaps (PBG) are disallowed bands of wavelengths. The presence of point defect, line defect, or both discontinues the periodicity of this band gap and localizes the propagation of light at these defect regions in the photonic crystal.

The use of photonic crystals (PhCs) is rapidly developing by the scientific and research communities since 1987. Researchers all around the world have reported many photonic crystal-based devices such as demultiplexer [3, 4], multiplexer [5], photonic crystal flat lens [6], Mach–Zehnder interferometer [7], optical switch [8, 9], optical logic gates [10], channel add drop filter [11], photonic crystal power splitters [12–16], etc.

The key building block is beam splitters in photonic multifunctional devices and systems. There are mainly three distinct ways to split the power of an incoming signal equally into some output ports using a T junction or Y junction [12], using a directional coupler [13, 14] and using a photonic crystal ring resonator [15, 16]. The size of device becomes large with great energy losses if we use Y junction or T junction beam splitter. Ideally, the input power should be divided equally into some output ports by a splitter without significant radiation losses or reflection. Practically, the complete transmission is not possible in conventional devices due to considerable reflection.

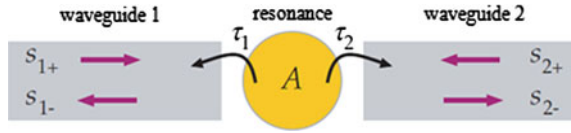
In this paper, the design of 1×6 power splitter based on resonance of the ring resonators is proposed and the coupling characteristics between ring resonator and waveguide are also presented. Previously, the researchers designed 1×2 power splitters [15] and 1×4 power splitters [16] based on the photonic crystal ring resonator (PCRR). The theoretical and numerical analyses of such structures are studied in this paper. The OptiFDTD simulation software of Optiwave System Inc. (using official license) is used to design, simulate, and analyze this 2D PhC structure.

The paper is organized as follows: Sect. 2 describes CMT, In Sect. 3 design parameters of 1×6 splitter are discussed, Sect. 4 presents FDTD simulation and analysis, and Sect. 5 concludes the future possibilities to improve the efficiency with negligible losses of this T-shaped 1×6 splitter.

2 Coupled Mode Theory

The theoretical analysis of the coupling of a cavity resonator to the waveguide system is called coupled mode theory (CMT) [17]. This time-dependent CMT applies to the cavity bend and splitter. The temporal coupled-mode equations describe the balance between incoming and outgoing field fluxes. There are two types of losses inside the cavity which are intrinsic losses and coupling to the waveguides losses. The energy leakage inside the cavity into the structure due to surrounding the cavity and coupling to the waveguide are responsible by the intrinsic quality factor ($Q_0 = 1/\tau_0$) and the external coupling factor ($Q_e = 1/\tau_e$),

Fig. 1 The abstract diagram of resonant cavity connected to two single-mode waveguides



respectively. The value for Q_e is different for each cavity mode because it is dependent on the symmetry of the cavity mode with respect to the waveguide modes.

The temporal coupled-mode theory describes the structure as a resonant cavity connected to two single-mode waveguides as shown in Fig. 1 [17]. The cavity mode has some resonant frequency ω_0 and decays with lifetime τ_1 and τ_2 into the two waveguides. The condition for 100 % transmission on resonance is $\tau_1 = \tau_2$ by symmetry. We assume that there is weak coupling between the various elements in temporal coupled-mode theory.

The equations of coupling of the cavity to the waveguides in terms of the field amplitudes in those components are derived. There are some assumptions such as weak coupling, linearity, time invariance, conservation of energy, and time-reversal invariance. The weak coupling is the most important assumption. Let the fields in the cavity are proportional to some variable A and the electromagnetic energy stored in the cavity is $|A|^2$. The fields in the waveguide are equal to the sum of incoming and outgoing waveguide modes. There is no incident power from the waveguides so begin with the cavity mode itself. We assume that the mode will decay exponentially over time with some lifetime τ due to weak coupling. If the cavity has two loss mechanisms, with decay constants τ_1 and τ_2 , then the net lifetime is given by $1/\tau = 1/\tau_1 + 1/\tau_2$. The amplitude inside the cavity is shown in Eq. (1) [17].

$$A(t) = A(0)e^{-i\omega_0 t - t/\tau} \quad (1)$$

The differentiation of Eq. (1) with respect to time is given below in Eq. (2) [17].

$$\frac{dA}{dt} = -i\omega_0 A - \frac{A}{\tau} \quad (2)$$

The most general equations (with assumptions) are given in Eq. (3) and Eq. (4) [17].

$$\frac{dA}{dt} = -i\omega_0 A - \frac{A}{\tau_1} - \frac{A}{\tau_2} + k_1 S_{1+} + k_2 S_{2+} \quad (3)$$

$$S_{l-} = \beta_l S_{l+} + \gamma_l A \quad (4)$$

where

- S_{l+} the amplitude of the mode going toward the cavity in the waveguide l
- S_{l-} the amplitude of the mode going away from the cavity in the waveguide l
- β_l reflection coefficient

k_l and γ_l the coupling strength of the cavity with respect to the waveguide

The constant γ_l is calculated using the conservation of energy. Consider the case where $\tau_2 \rightarrow \infty$, $S_{1+} = S_{2+} = 0$, then the energy $|A|^2$ is decreasing exponentially as $A(t) = A(0)e^{-i\omega_0 t - t/\tau}$. This energy going into the outgoing power $|s_{-1}|^2$ is shown by Eq. (5) [17].

$$-\frac{d|A|^2}{dt} = \frac{2}{\tau_1} |A|^2 = |s_{1-}|^2 = |\gamma_1|^2 |A|^2 \quad (5)$$

Therefore, $\gamma_1 = \sqrt{2/\tau_1}$ is calculated using Eq. (5). Similarly, we find $\gamma_2 = \sqrt{2/\tau_2}$ by putting $\tau_1 \rightarrow \infty$. The constants k_l and β_l are obtained by time-reversal symmetry. We get the values $\beta_l = -1$ and $k_l = \gamma_l = \sqrt{2/\tau_l}$. Finally, the temporal coupled-mode equations for two port system in Fig. 1 are given in Eq. (6) [17] and Eq. (7) [17].

$$\frac{dA}{dt} = -i\omega_0 A - \sum_{l=1}^2 \frac{A}{\tau_l} + \sum_{l=1}^2 \sqrt{\frac{2}{\tau_l}} S_{l+} \quad (6)$$

$$S_{l-} = -S_{l+} + \sqrt{\frac{2}{\tau_l}} A \quad (7)$$

The temporal coupled-mode equations for the T-shaped splitter of three-port system with assuming junction as weak resonance is shown in Eq. (8) [17] after modifying the Eq. (6).

$$\frac{dA}{dt} = -i\omega_0 A - \sum_{l=1}^3 \frac{A}{\tau_l} + \sum_{l=1}^3 \sqrt{\frac{2}{\tau_l}} S_{l+} \quad (8)$$

The reflection and the transmission spectra for three port system of splitter are calculated using Eqs. (7) and (8) with $S_{2+} = S_{3+} = 0$. Equation (9) [17] represents the back of reflection into waveguide 1. Equation (10) [17] and Eq. (11) [17] represents the transmission into waveguide 2 and waveguide 3, respectively, for three port system of T-shaped splitter.

$$R(\omega) = \frac{|S_{1-}|^2}{|S_{1+}|^2} = \frac{(\omega - \omega_0)^2 + \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} - \frac{1}{\tau_3}\right)^2}{(\omega - \omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right)^2} \quad (9)$$

$$T_{1 \rightarrow 2}(\omega) = \frac{|S_{2-}|^2}{|S_{1+}|^2} = \frac{\frac{4}{\tau_1 \tau_2}}{(\omega - \omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right)^2} \quad (10)$$

$$T_{1 \rightarrow 3}(\omega) = \frac{|S_{3-}|^2}{|S_{1+}|^2} = \frac{\frac{4}{\tau_1 \tau_3}}{(\omega - \omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right)^2} \quad (11)$$

The reflection is zero and the transmission is 100 % from waveguide 1 to waveguide 2 and 3 at $\omega = \omega_0$, if the decay rate must satisfy the condition which is shown by Eq. (12) [17].

$$\frac{1}{\tau_1} = \frac{1}{\tau_2} + \frac{1}{\tau_3} \quad (12)$$

3 Design Parameters

The design of two-dimensional PhC T-shaped 1×6 power splitter with 2×2 ring resonator consists of square lattice as shown in Fig. 2. The dielectric rods of GaAs ($n = 3.46$) with radius of $0.185a$ are embedded in air where a and n are the lattice constant and refractive index, respectively. The value of the lattice constant is 540 nm. The 2D FDTD method and PWE method are used to calculate the spectrum of power transmission and the band diagram, respectively. The perfectly matched layers (PMLs) absorbing boundary conditions are used at the boundary of the computational region to absorb the reflections from the outer boundary. The number of PMLs is set to be 12. The time-varying electric and magnetic fields are measured by a detector inside waveguide channel. The size of wafer is $22.5 \mu\text{m} \times 14.3 \mu\text{m}$ for

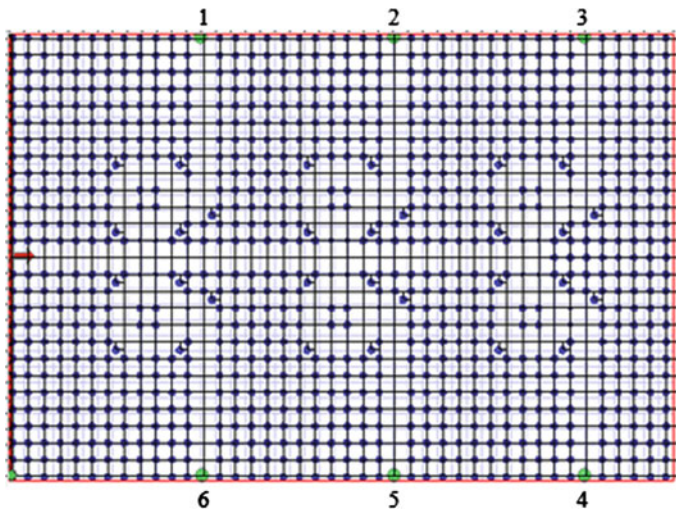


Fig. 2 Design of 1×6 power splitter of T-shaped based on 2×2 ring resonator

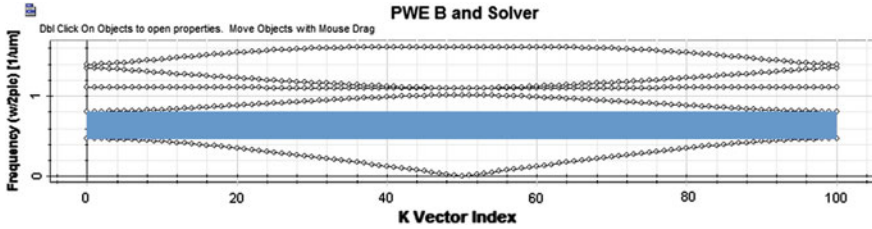


Fig. 3 Band diagram of square photonic crystal lattice

1×6 power splitter. The number of rods is 27×42 in x - z plane. There is one input port, six output ports, and six PCRRs. The single photonic crystal ring resonator is placed between each input and output ports. Four scatter rods are placed at the four corners of each PCRR. The scatter rods with half lattice constant were added at each corner to obtain uniform transmission. Six observation points are placed at the six output ports and these six observation points are labeled 1, 2, 3, 4, 5, and 6 as shown in Fig. 2.

There is coupling between the waveguide and ring resonator by placing the ring resonator near the waveguide. The electromagnetic energy in one waveguide is transferred to the other waveguide through the ring resonator. There must be low reflection, low loss, and broad bandwidth while designing a bend. The place of resonant structure follows T junction and 90° bends with respect to the waveguide intersection. The bends of 90° are placed back to back in the opposite direction of waveguide intersection.

The change in the size of ring resonator affects the whole spectral characteristic of splitter. This means that the numerical simulation is affected by the various parameters of ring such as size, and position of the ring. The proper choice of parameters is necessary for high transmission with low loss and reflection. The radius of scatter rods (R_s) and coupler rods (R_c) are optimization parameters to obtain high transmission. After an optimization process, we obtain high transmission by selecting the values of $R_c = 0.5r$ and $R_s = 1.1r$ for T-shaped splitter.

The wide band gap in the range of $0.470897 \leq 1/\lambda \leq 0.806754$ for TE mode is calculated by PWE method [18] where λ is the wavelength in free space. The corresponding wavelength range extends from 1.2395 to 2.1236 μm . The structure of band diagram is shown in Fig. 3.

4 Numerical Simulation

The vertical input plane is used to inject a Gaussian-modulated continuous wave signal at the wavelength of 1580 nm. Six observation points are used to calculate the transmission at six output ports. The FDTD method is used to calculate the transmission spectrum by running 2D 32 bit simulation parameters with 5000 time

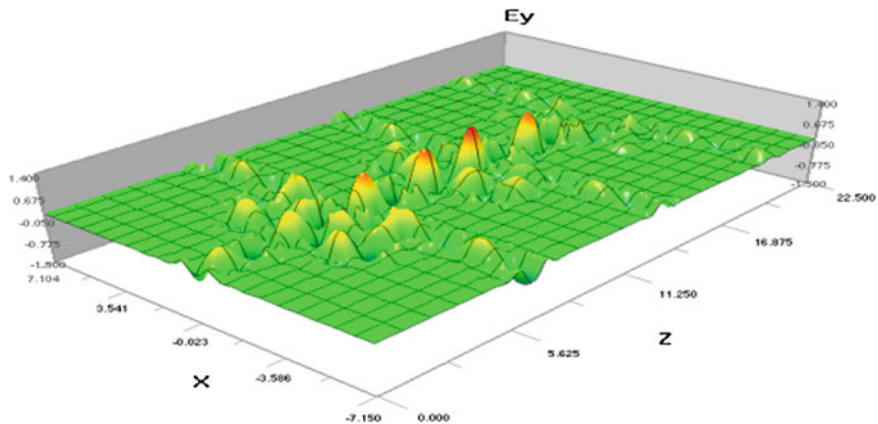


Fig. 4 Snapshot of simulation of the splitter

steps. The snapshot of simulation of T-shaped 1×6 splitter using OptiFDTD software at $\lambda = 1580$ nm is shown in Fig. 4. The electric field pattern of T-shaped 1×6 splitter at $\lambda = 1580$ nm is symmetrical due to the symmetrical structure on the direction of input waveguide. The intensity of electric field at output ports 1, 2, and 3 is equal to the output ports 6, 5, and 4, respectively, due to the symmetry of the structure.

The FDTD analyzer is used to view the output response of the 1×6 power splitter. Frequency discrete Fourier transform (DFT) is used to obtain the transmission spectra of the splitter. The DFT of E_y field is analyzed by selecting the observation points. The transmission spectrum of the PhC 1×6 splitter is shown in Fig. 5. The analysis of 1×6 power splitter is done by varying the radius of scatter rods and the radius of coupler rods. The value of the radius of scatter rods ($R_s = 1.1r$) and the radius of coupler rods ($R_c = 0.5r$) is determined through an optimization process. The power transmission at output ports 1, 2, 3 are exact the same as 6, 5, and 4, respectively, at 1580 nm wavelength due to equal energy flow on both sides of input waveguide.

5 Conclusion

In this paper, the design of the PCRR-based T-shaped splitter with square lattice is investigated by FDTD method and analytically by CMT. The PWE method is used to calculate the PBG. It has been observed that the transmission efficiency of the 1×6 splitter is dependent on the radius of the scatter rods and the radius of coupling rods. The optimization technique is used to get the values of the radius of the scatter rods, $R_s = 1.1r$, and the radius of coupling rods, $R_c = 0.5r$. The overall size of the chip is around $22.5 \mu\text{m} \times 14.3 \mu\text{m}$. Further optimization for this 2D

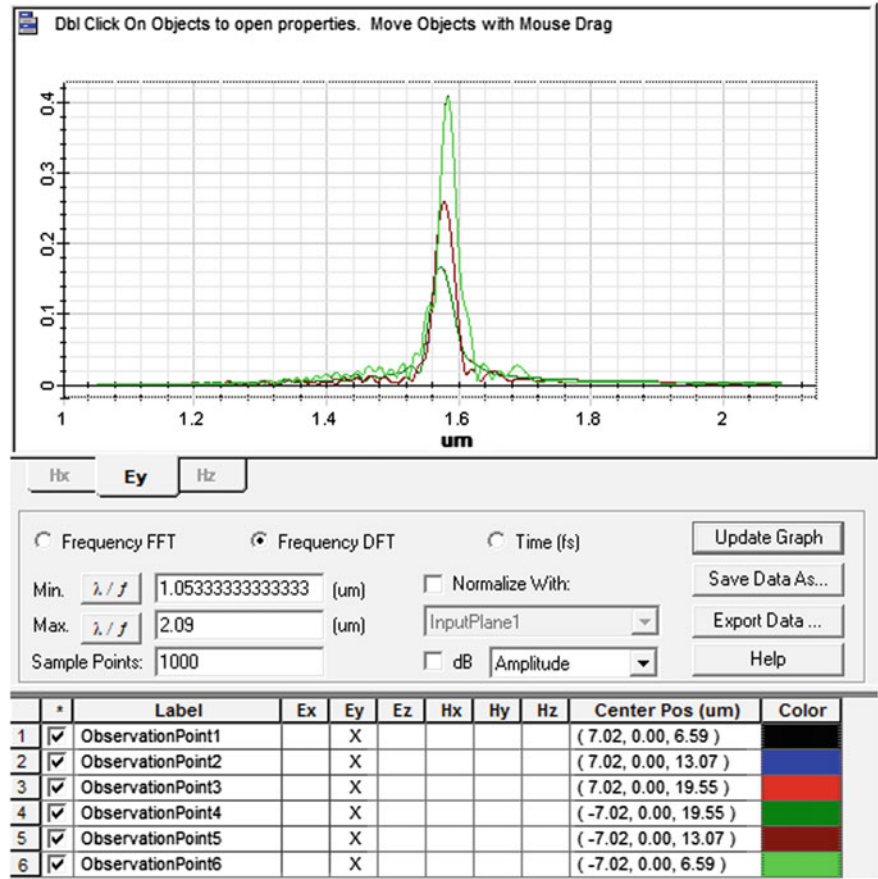


Fig. 5 The transmission spectrum of the photonic crystal 1 × 6 power splitter

photonic crystal 1 × 6 splitter based on the ring resonator remains for future work to get the better response. The output response of T-shaped 1 × 6 splitters based on the 2 × 2 ring resonator shows higher transmission efficiency, ultracompact size ($322 \mu\text{m}^2$), and better splitting ratio with $R_c = 0.5r$ and $R_s = 1.1r$ at the wavelength of 1580 nm compared to the previous works. Hence, such kind of devices may be useful in photonic integrated circuits, optical networking communication and applications, etc.

References

1. Eli. Yablonovitch: Inhibited spontaneous emission on solid-state physics and electronics. Phys. Rev. Lett., vol. 58 (20): 2059–2062 (1987).

2. S. John: Strong localization of phonics in certain disordered dielectric Super-lattices. *Phys. Rev. Lett.* 58: 2486–2489 (1987).
3. Hadi Ghorbanpour, Somaye Makouei: 2-channel all optical demultiplexer based on photonic crystal ring resonator, *Frontiers of Optoelectronics*. Volume 6, Issue 2: 224–227(2013).
4. Hui Liu, Xiang-bao Cai: Study of wavelength demultiplexer based on two-dimensional photonic crystals. *Optoelectronics Letters*, Volume 4, Issue 5: 339–341(2008).
5. G. Manzacca, D. Paciotti, A. Marchese, M. S. Moreolo, G. Cincotti: 2D photonic crystal cavity-based WDM multiplexer, *Photonics and Nanostructures - Fundamentals and Applications*. vol. 5: 164–170 (2007).
6. W. S'migaj, B. Gralak, R. Pierre, G. Tayeb: Antireflection gratings for a photonic crystal flat lens. *Opt. Lett.* 34 3532–3534 (2009).
7. Y. Geng, X. Li, X. Tan, Y. Deng, Y. Yu: A cascaded photonic crystal fiber Mach–Zehnder interferometer formed by extra electric arc discharges. *Applied Physics B*, Volume 102, Issue 3: 595–599 (2011).
8. A.T. Rahmati, N. Granpayeh: Kerr nonlinear switch based on ultra-compact photonic crystal directional coupler. *Optik* 122 502–505 (2011).
9. H.Z. Wang, W.M. Zhou, J.P. Zheng: A 2D rods-in-air square-lattice photonic crystal optical switch. *Optik* 121 1988–1993 (2010).
10. X. Susan Christina, A. P. Kabilan: Design of optical logic gates using self-collimated beams in 2D photonic crystals, *Photonic Sensors*. vol 2, Issue 2: 173–179 (2012).
11. W. ChunChih, C. LienWen: Channel drop filters with folded directional couplers in two-dimensional photonic crystals. *Phys. B* 405 1210–1215 (2010).
12. A. Ghaffari, F. Monifi, M. Djavid, M.S. Abrishamian: Analysis of photonic crystal power splitters with different configurations. *J. Appl. Sci.* 8 1416–1425 (2008).
13. Park, H.S. Lee, H.J. Kim, K.M. Moon, S.G.B.H. Lee, S.G.O. Park, E.H. Lee: Photonic crystal power-splitter based on directional coupling. *Opt. Express* 12 3599–3604 (2004).
14. Y. Tianbao, W. Minghua, J. Xiaoqing, L. Qinghua, Y. Jianyi: Ultracompact and wideband power splitter based on triple photonic crystal waveguides directional coupler. *Opt. A: Pure Appl. Opt.* 9 37–42 (2007).
15. Ghaffari, F. Monifi, M. Djavid, M.S. Abrishamian: Photonic crystal bends and power splitters based on ring resonators. *Opt. Commun.* 281 5929–5934 (2008).
16. HU Ping, GUO Hao, LIAO Qinghua: Large separating angle multiway beam splitter based on photonic crystal ring resonator. *SPIE-OSA-IEEE Vol.* 7631, 76312C-1(2009).
17. John D. Joannopoulos, Steven G. Johnson, Joshua N. Winn, Robert D. Meade: *Photonic crystal molding the flow of light*. second edition (2008).
18. K.M. Leung and Y.F. Liu: Photon band structures: The plane-wave method. *Phys. Rev. B*, vol. 41: 10188–10190 (1990).

Performance Evaluation of Vehicular Ad Hoc Network Using SUMO and NS2

Prashant Panse, Tarun Shrimali and Meenu Dave

Abstract In current scenario each and every person is anxious about security and privacy. Vehicular correspondences frameworks have ways to deal with give well-being measures and solace to drivers. Vehicular communication is based on wireless short-range technology that enables impulsive information interchange among vehicles and with roadside stations. A new type of network called vehicular ad hoc network (VANET) is available for providing alerts to the vehicles on highways. VANET is vehicular ad hoc network, in which mobile nodes are replaced by vehicles. Vehicular network is used to alert a driver so that accidents can be reduced and also avoid congestion on highways. This can be used for postaccident investigation as well. Frequently changing environment of VANET leads to various challenges. In this paper, the performance of vehicular ad hoc network is evaluated by focusing several key factors and reactive routing strategy.

Keywords VANET · Ad hoc network · D2ITS · ITS · AODV · Ultrasonic sensor · Roadside unit · SUMO · NS2

1 Introduction

In current years, three revolutions have been seen in vehicle development. It includes stronger engines, safety features, and most important is accident prevention using new technologies. Nowadays accident avoidance and prevention systems are used which is active and also help vehicles itself and drivers to avoid accidents.

Prashant Panse (✉) · Meenu Dave
JaganNath University, Jaipur, India
e-mail: prashantpanse@gmail.com

Meenu Dave
e-mail: meenu.s.dave@gmail.com

Tarun Shrimali
Career Point Technical Campus, Rajsamand, India
e-mail: shrimalitarun@gmail.com

There are some passive safety devices available with active devices such as airbags, head restraints to reduce the severity of an accident. In today's scenario accident prevention systems are available which are based on V2V, V2I communication. These system names include electronic brake force distribution, infrared night vision systems. Following is a detail of examples of prevention system for accidents.

1.1 Driver Fatigue Monitoring

Due to exhaustion and fatigue, it may be possible that driver falls asleep while driving which results in accidents. The system developed to help drivers is a driver monitoring system which activates the autonomous emergency braking when the recorded eye movement is mismatched to the routine eye motion. To do so, a sensor is embedded in eye gear to monitor eye movement. A threshold level of mismatch is set, exceeding of which causes an alarm sound to alert the driver [1].

1.2 Blind Spot Accident Prevention System Based on Sensors

In this system, when the obstacle or bystander is detected by the blind spot detection device, the device triggers a first level alarm. A second level of visual and capable of being heard alert is activated if the vicinity of hindrance is distinguished even after a period deferral of first level caution. The second level caution alarms the system administrator (operator) of the unsafe circumstance and the vehicle will stop naturally [1].

2 Evaluation of Accident Prevention Systems

It is beneficial to road safety by reducing accident numbers and the severity of accidents using advanced accident prevention systems. Also, there are a number of advantages for transport operators such as less vehicle downtime and lower insurance premiums. It is observed against most of the systems, that the main cause of an accident involving heavy commercial vehicles is not effectively targeted. The main causes of accidents according to ETAC study are: not respecting intersection rules, use of improper maneuver when changing lanes and nonadapted speed [1]. There may be considerable impact on driver vehicle communication due to non-coordination in development of various system, is another area of concern. In fact, driver may start ignoring warning signals if it occurs regularly. Similarly,

transport operator's choice should be taken into account so as the prevention system are acceptable to them, as experience shows that these systems are not appreciated. Advance accident prevention system may contribute to false sense of safety, by which irresponsible driving offsets the safety benefits of the system. At last, research on accident prevention and piloting of technology is often used as a backdoor to the influencing and development of technical legislation, meaning there is a clear lack of transparency in the drafting of legislation.

3 Architecture of VANET

Vehicular ad hoc network is used for communication and cooperative driving between cars [1]. Vehicle-to-Vehicle (V2V) correspondence permits sharing the remote channel for versatile applications to plan the routes, controlling movement clogging, or activity well-being change, e.g., maintaining a strategic distance from accident circumstances [2]. For providing so as to diminish the quantity of lethal roadway mishaps early notices rising remote advances for V2V and V2R correspondence, for example, dedicated short-range communication (DSRC), seems quite encouraging. [3]. Broadcast method is frequently used in inter-vehicular communication (IVC) system.

Remote access in vehicular situations, characterizes structural planning for Astute Transportation Frameworks has received 802.11p, which is an amendment in 802.11 standard of IEEE for vehicular interchanges [2–4]. Figure 1 depicts the communication between multiple vehicles.

The development in wireless technologies has permitted researchers to devise communication systems where vehicles take part in the communication. On the off chance that vehicles can specifically correspond with one another and with foundation, an altogether new worldview for vehicle well-being applications can be

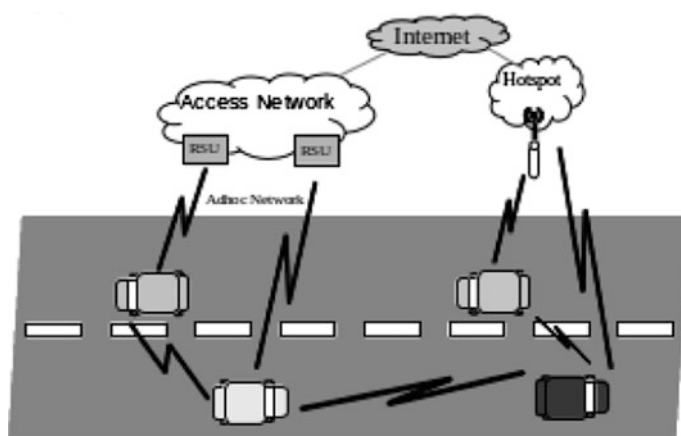


Fig. 1 Vehicle-to-vehicle reference architecture

made [5]. It also provides traffic alerts and information on time about jam; accidents on highway, increase road safety, and at the same time improve safe driving. Safety applications can be partitioned into active, passive, and proactive category [6]. Vehicular network permits correspondence among close-by vehicles and in the middle of vehicles and adjacent settled roadside equipment [7].

4 Simulation

Vehicular ad hoc network is simulated using SUMO and NS2 simulator. SUMO is an open-source traffic simulation tool. SUMO network consists of junction, edges, and nodes. SUMO network consists of node files (.nod.xml), edge file (.edg.xml), route file (.rou.xml), network file (.net.xml), and configuration file (.sumo.cfg.xml). In this paper road network is created using SUMO then we created traffic on this. Further, SUMO configuration is converted into tcl file and simulation is done in NS2. The sample file of implementation is shown below.

Sample of new.nod.xml

```
<nodes>
<node id="node0" x="100.0" y="300" type="priority"/>
<node id="node1" x="500" y="300" type="traffic_light"/>
<node id="node2" x="100.0" y="600" type="priority"/>
</nodes>
```

Sample of new_EDGE.edg.xml

```
<edges>
<edge id="edgeS-0-1" fromnode="node0" tonode="node1"
priority="75"
nolanes="3" speed="40" />
</edges>
```

Sample of new12.net.xml

```
<routes>
<vehicle id="flow0_0" depart="0.00">
<route edges="edgeS-0-1 edgeS-1-0 edgeS-0-1 edgeL-1-4
edgeL-4-1 edgeS-1-0 edgeL-0-2 edgeL-2-0 edgeS-0-1 edgeL-
1-5 edgeL-5-7 edgeL-7-9 edgeL-9-7 edgeL-7-8 edgeL-8-7
edgeL-7-10 edgeL-10-7 edgeL-7-8 edgeL-8-7 edgeL-7-9
edgeL-9-7 edgeL-7-10 edgeL-10-7 edgeL-7-5 edgeL-5-6
edgeL-6-5 edgeL-5-6 edgeL-6-5 edgeL-5-1 edgeS-1-0 edgeL-
0-2 edgeL-2-0 edgeS-0-1 edgeL-1-3 edgeL-3-1 edgeL-1-3
edgeL-3-1 edgeL-1-5 edgeL-5-7 edgeL-7-9 edgeL-9-7"/>
</vehicle>
</routes>
```

Sample of SUMO configuration file new12.sumo.cfg

```
<configuration>
<input><net-
filevalue="/home/mitm/Desktop/PhD/new12.net.xml"/>
<route-files
value="/home/mitm/Desktop/PhD/new12.net.xml.rou.xml"/>
<additional-files value=""/>
<junction-files value=""/>
</input>
</configuration>
```

This configuration gives us result which shown Fig. 2.

4.1 Simulation Parameters

VANET is simulated considering various network parameters which are tabulated in Table 1.

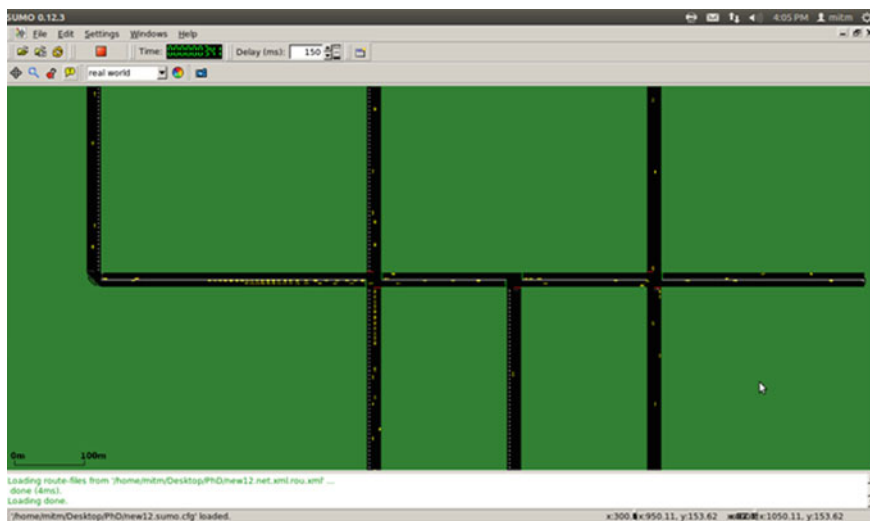


Fig. 2 SUMO scenario

Table 1 Simulation parameters

Parameters name	Parameter values
Number of nodes	20, 40, 60, 80, 100
Simulation time	985 s
Traffic type	CBR
Connection type	UDP
Routing protocol	AODV
Queue type	DropTail

4.2 Simulation Scenario

VANET is simulated considering different number of nodes such as 20, 40, 60, 80, and 100. VANET simulation scenario with 20 nodes is shown in Fig. 3.

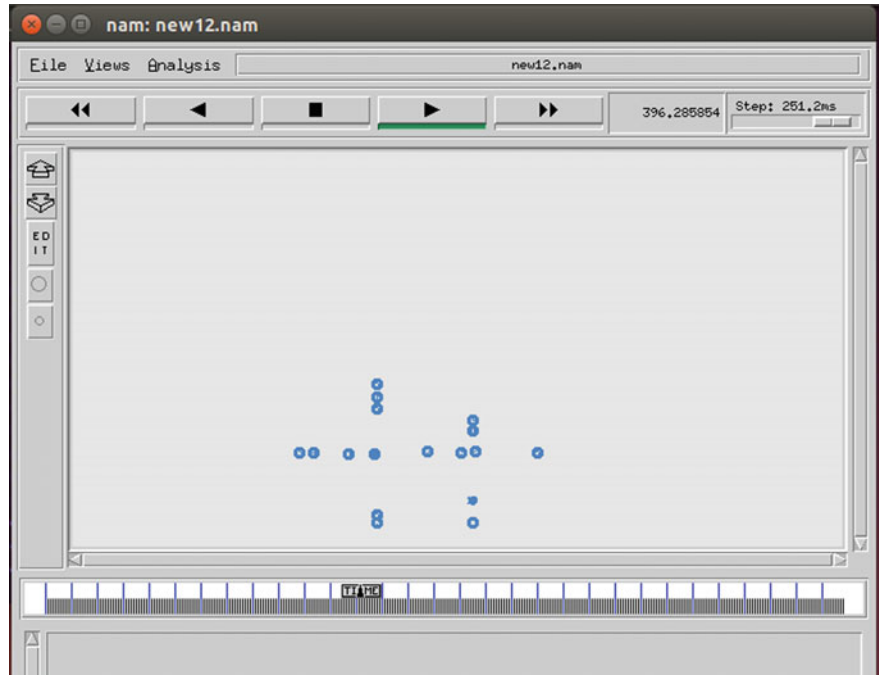


Fig. 3 Simulation scenario of 20 nodes in NAM

5 Result Analysis

The performance of vehicular ad hoc network is evaluated considering distinctive parameters, for example, data delivery rate, throughput, routing overhead, average end-to-end delay, and remaining energy, which are computed on the basis of simulation.

Throughput is defined as average number of bits, bytes, or packets per unit time (Fig. 4).

Data delivery rate is the ratio of received packet and sum of dropped and received packets in a network (Fig. 5).

Fig. 4 Throughput versus number of nodes

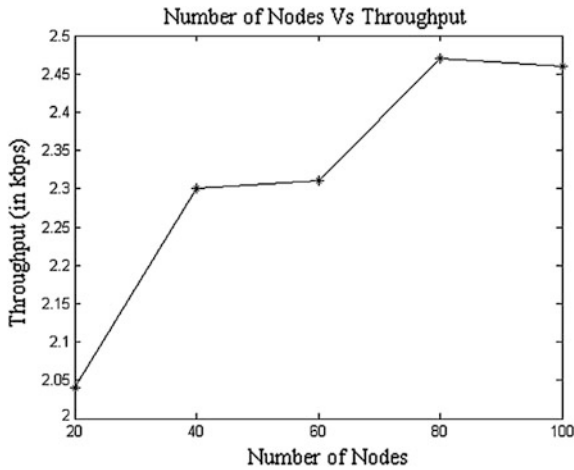


Fig. 5 PDR versus number of nodes

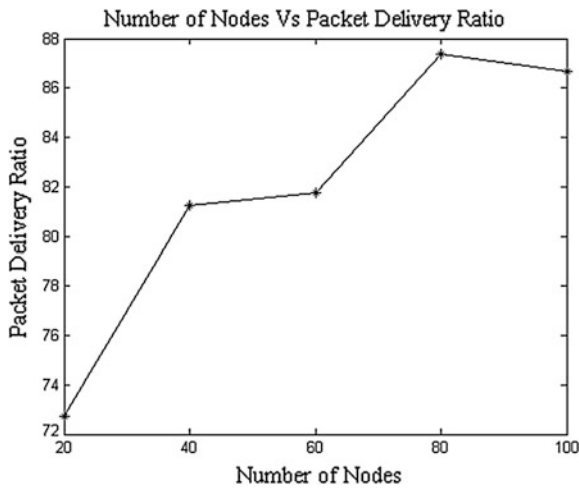


Fig. 6 End-to-end delay versus number of nodes

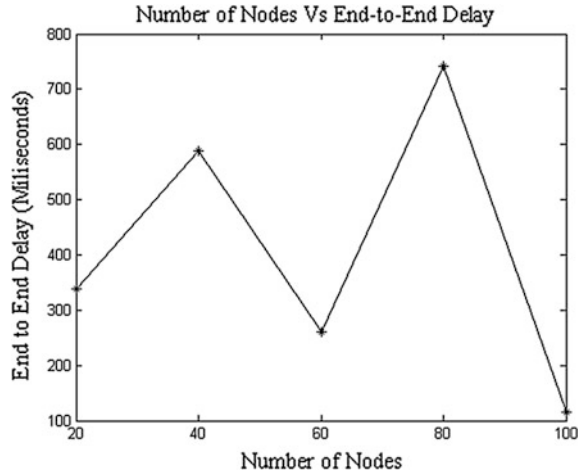
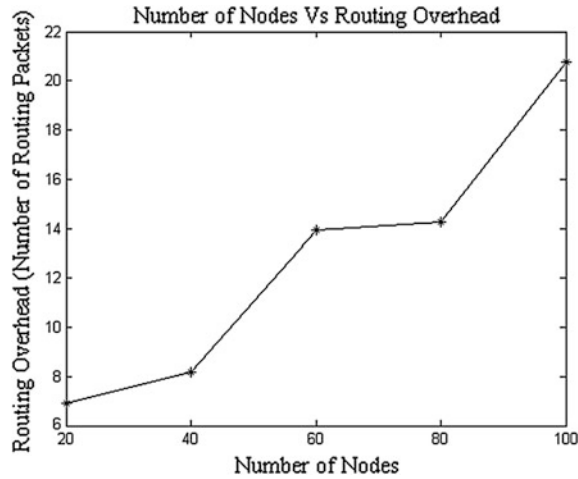
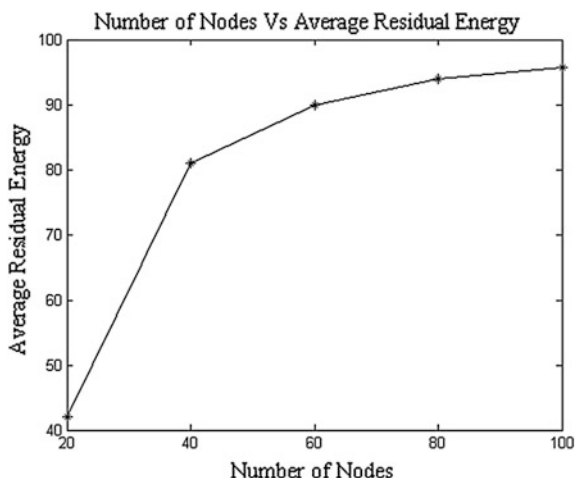


Fig. 7 Routing overheads versus number of nodes



End-to-end delay is the time required by a packet to reach its destination (Fig. 6). Routing overhead is the total number of routing packets traversed in network over simulation time (Fig. 7). Average remain energy is the average residual energy of network over simulation time (Fig. 8).

Fig. 8 Average remain energy versus number of nodes



6 Conclusion

Vehicular ad hoc network is simulated with different parameters. First, we created a SUMO network and a scenario for different number of vehicles, and then it is converted into tcl script using MOVE. We also evaluated performance of network based on throughput, end-to-end delay, average residual energy, packet delivery ratio, routing overhead using AODV protocol in NS2.

References

1. IRU Position on Advanced Accident Prevention Systems. International Road Transport Union, Switzerland (2009).
2. Padmavathi K., Maneendhar R.: A Surveying on Road Safety Using Vehicular Communication Networks, *Journal of Computer Applications*, ISSN: 0974-1925, Volume-5, Issue EICA2012-4, 460–465 (2012).
3. Karp, B., Kung, H.T.: GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking (MobiCom '00)*. ACM, New York, NY, USA, 243–254 (2000).
4. Füller, H., Mauve, M., Hartenstein, H., Käsemann, M., Vollmer, D.: Location-Based Routing for Vehicular Ad-Hoc Networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, Volume 7 Issue 1, 47–49 (2003).
5. Hartenstein, H., Laberteaux, K.P.: *VANET Vehicular Applications and Inter-Networking Technologies*. Wiley (2010).
6. Zeletin, R.P., Radusch, I., Rigani, M.A.: *Vehicular-2-XCommunication*, Springer-Verlag Berlin Heidelberg (2010).
7. Moustafa, H., Zhang, Y.: *Vehicular Networks Techniques, Standards, and Applications*, Auerbach (2009).

An Intrusion Detection System for Detecting Denial-of-Service Attack in Cloud Using Artificial Bee Colony

Shalki Sharma, Anshul Gupta and Sanjay Agrawal

Abstract Cloud computing is a technology which allows users to share resources and data over the Internet. Cloud computing represents the maturing of technology and is a pliable, cost-effective platform which provides business/IT services over the Internet. Although there are various benefits of adopting this technology, there are some significant barriers to it and one of them is security. Cloud computing is still growing and there is still uncertainty about how security is achieved, at all levels (network, host, application, and data), in cloud. In computing environment like cloud where whole infrastructure is shared by millions of users, attacks like denial-of-service are likely to have a much greater footprint than other attacks. The main aim of denial-of-service attack is the disruption of services by attempting to limit access to a machine or service instead of subverting the service itself. This paper tested the efficiency of artificial bee colony, a swarm approach, for finding denial-of-service attack in a cloud environment and finds that it is useful in tackling denial-of-service attacks.

Keywords Cloud computing · Denial-of-service · Artificial bee colony

1 Introduction

As the field of cloud computing is growing so are the security issues pertaining to it. The world cloud is very appealing as it provides the user with a lot of resources at one place without much of effort. Cloud provides its user with (1) on demand

Shalki Sharma (✉) · Sanjay Agrawal
NITTTR, Bhopal, India
e-mail: shalkisharma27@gmail.com

Sanjay Agrawal
e-mail: sagrawal@nitttrbpl.ac.in

Anshul Gupta
MPSTME, NMIMS, Mumbai, India
e-mail: anshul.gupta@niims.edu

self-service (2) broader network access (3) resource pooling (4) rapid elasticity (5) measured services [1]. There are various benefits of using cloud over a traditional approach such as cloud helps to reduce the cost; it provides global access, unlimited storage capacity, improved performance, and many more. But with this attraction comes a severe issue of the security in cloud. Threats like man-in-the-middle attack, denial-of-service attack are always present and attackers have become more prominent and active in using these kinds of attacks for disrupting the services of cloud and making them unavailable to the intended users. Cloud Security Alliance [2] has defined (1) data breaches (2) data loss (3) account hijacking (4) insecure APIs (5) denial-of-service (6) malicious insiders (7) abuse of cloud services (8) insufficient due diligence (9) shared technology issues as “Notorious Nine,” nine critical threats to cloud computing. Hackers in the past have tried to attack and some have been successful also. On August 6, 2009, twitter went down abruptly for 2 h and the reason for this shut down was denial-of-service attack [3, 4].

With the advancement made in cloud, security has become an important aspect both with respect to the user and as well as to the CSP. In our proposed work, we are using artificial bee colony (ABC) technique for the detection of denial-of-service attack in a cloud environment. The rest of the paper is sub partitioned into four sections. In Sect. 2, a brief definition of DoS attack and its detection approaches are provided. In Sect. 3, proposed methodology is summarized. Section 4, discusses the results and we conclude the paper with Sect. 5.

2 Denial-of-Service Attack

Denial-of-service is an attack that attempts to make the resources or services unavailable to the users for infinite amount of time by flooding it with useless traffic [5, 6]. Numerous approaches have been proposed in the past for detecting these kinds of attacks. Some of the techniques and related work are mentioned below.

2.1 *Related Work*

2.1.1 **Malicious Detection**

Mahajan Pushback approach [7] uses two techniques; aggregate congestion control (ACC) and pushback. Local ACC uses identification algorithm for finding the cause of congestion and control algorithm for reducing its effect. Second mechanism, Pushback, allows router to request their adjacent upstream router to rate limit the specified aggregate. Crowding at router level is detected by Local ACC and devices a congestion signature and translates into router filter. Network traffic and high bandwidth aggregate are defined by signature and local ACC defines a rate limit for

this aggregate. This rate limit is propagated to the intermediate upstream neighbors, by Pushback that contributes to the largest amount of traffic.

Lo et al. [8] proposed the use of distributed IDS and of cooperative defense for each of the cloud by the IDS. Each cloud is provided with its own IDS and alerts are generated by the IDS who are under the attack. Trustworthiness of the alert is defined by judgment criteria. Block tables are used to keep track of all the alerts generated and if any new alert is found, it is added in the table, thus helping in an early detection. Alerts so generated are categorized among serious, moderate, and slight; depending upon the type of the attack. The overall benefit is that it forestalls the entire system from a single point failure.

Approach proposed by Lua et al. [9] aims to detect DoS attack using intelligent fast flux swarm network. Fast flux technique maintains connectivity among swarm nodes, clients, and servers. To maintain parallel and distributed optimization IWD was used. Swarm network was built on two concepts: fast flux technique in DNS and organization of swarm. Client reaches the server via fully qualified domain name and the request is forwarded to the designated server via community exit node. Results are reverted back to the client through the swarm network. Swarm network reconfigures itself constantly using IWD as it is highly resistant to sudden changes in network. The proposed approach is highly robust in nature.

Anitha et al. [10] proposed the use of packet marking approach for the detection of DoS attack. CLASSIE, rule set-based detection, was used to discriminate between legitimate and illegitimate attacks. The proposed method was checked by HX-DOS attacks cloud web services. CLASSIE is situated one hop away from the host. Whenever an HX-DOS attack is detected, CLASSIE drops the packets and they are subjected to marking, done both on edge and core routers. RAD method allows incoming messages to pass or to drop and is situated one hop away from the victim. RAD also avoids spoofing. The technique reduces the false positive rate.

Joshi et al. [5] uses cloud trace back (CTB) for detecting DoS attacks. CTB uses SOA for tracing back the true source of the attack and is based on deterministic packet marking algorithm. CTB uses FDPM by integrating a cloud trace back mark (CTM) within the header of CTB. Back propagation neural network is used as Cloud Protector, to train and filter out the traffic. CTB removes the service provider's address by placing itself before the web server and hence all services are first sent to CTB. In case an attack has been observed, attacker will request CTB for the service and attacker will formulate a SOAP message. CTM is placed in the CTB header upon the receipt of this message and the message is forwarded to the web server. When an attack is observed mark is extracted and this will also filter out the attack traffic. If the attack was successful the victim will recover the CTM tag and thus revealing the true identity of the source.

Reddy et al. [11] proposed the use of quantum-inspired particle swarm optimization technique (QPSO) for the detection of DoS attack in a cloud. Anomaly-based detection was used for decision making. The technique was subdivided into two subphases training and testing. In training, normal traffic was trained using the quantum algorithm and in the testing phase abnormal traffic was

tested using the detection module of the algorithm. The observed were compared with QEA and the algorithm was found to be better than QEA.

3 Proposed Methodology

A lot of techniques and approaches have been proposed in the past for detection of these kinds of attacks. In our research, we are using artificial bee colony (ABC), a swarm approach [12], for detecting these kinds of attacks. In the proposed framework basic feature selection is done for each record, ABC working nature is determined and at the end we do decision making. For evaluating the accuracy of ABC, we are comparing it with QPSO and it was found that the average accuracy of ABC is better than QPSO.

3.1 Artificial Bee Colony (ABC)

ABC proposed by Karaboga, simulates the foraging behavior of honey bees [13]. The colony of honey bees consists of employed bees, onlookers, and scouts. Pseudo code of the algorithm is given below. The bee which is waiting on the dance area for making a decision to select the food source is the onlooker and the bee going to the food source, visited by it before, is the employed bee. Scouts are responsible for carrying out random search for finding new food source. The first half of the algorithm is of artificial employee bee and the second half of onlookers. Possible solutions to optimization problem are found by the position of the food source the nectar amount of a food source corresponds to the quality (fitness) of the associated solution, calculated by [13]. As the algorithm performs both global and local searches, it gives us efficient results.

$$\text{Fit}_i = \frac{1}{1 + f_i} \quad (1)$$

Steps of ABC Algorithm:

1. Start
2. Initialize the population
3. Employed bees finds a neighbor source for nectar and dances in hive
4. Each onlooker bee watches the dance chooses one of the neighbor sources depending on the dance.
5. Onlooker bee goes to that neighbor source and evaluate nectar amount.
6. Scouts replace abandoned food sources with new one
7. Determine best food source, so far
8. Repeat the steps from 3 to 7, until max is achieved
9. Stop

Probability, P_i , of choosing a new food source by onlooker bees is calculated by the following:

$$P_i = \frac{\text{Fit}_i}{\sum \text{Fit}_n} \quad (2)$$

where $n = 1 \dots \text{size of population}$

For finding the new solution, v_{ij} , in the neighborhood of old one, x_{ij} , following formula can be used:

$$V_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) \quad (3)$$

where k and j are randomly chosen indexes and Φ_{ij} is random function within the range $[-1, 1]$.

3.2 Dataset for Training and Testing

The efficiency of any bio-inspired network depends on the training data. The more accurate the training data is more is the performance of the network. Thus collection of data is critical factor for training and can be done in any of the three ways: using a real traffic, using sanitized traffic, or using a simulated traffic. Using simulated traffic is the most common and feasible way for obtaining data and for creating a test bed network and also for generating background traffic on the given network [5, 14]. Background traffic can be generated by employing complex traffic generators modeling actual network statistics or by employing a more simple traffic generator by fabricating smaller number of packets at a high rate. By adopting this approach, data can be distributed freely because there is no sensitive information in it and also assures that the generated traffic does not have any unknown attacks because simulator is producing this traffic. In our approach, on the whole we have generated the background traffic in CloudSim.

3.3 Framework

The proposed framework has been divided into three steps. First basic feature selection is done for each record. In this step basic network features are generated and traffic is recorded in a well-defined manner. The more detailed approach can be found in [15]. Second, we employ ABC algorithm and determine the working behavior of ABC and at the end we do decision making. Decision making is done using anomaly-based detection technique [16]. Anomaly-based technique

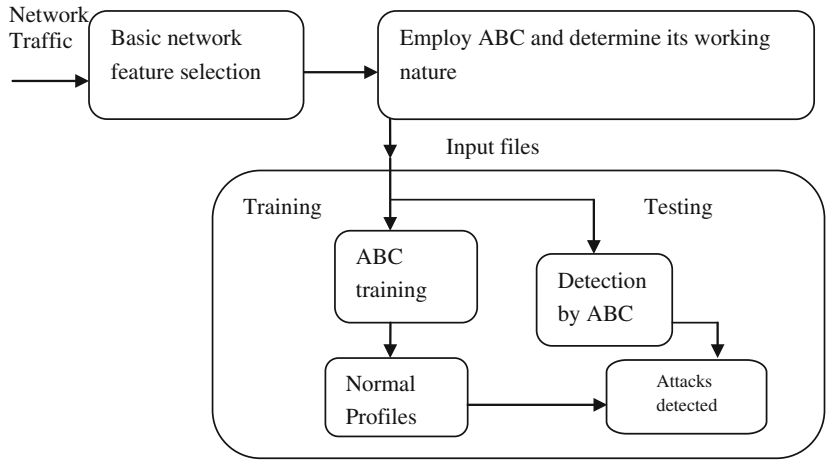


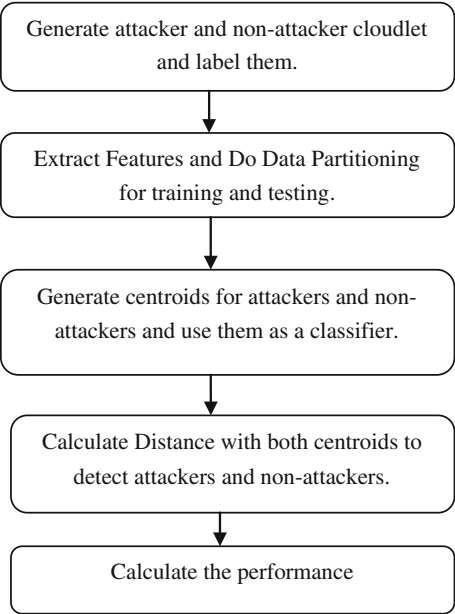
Fig. 1 Proposed framework

determines any kind of DoS attacks without having any kind of knowledge about the attacker. The technique is robust in nature as the attacker has to create a specific attack which appears as a normal traffic to the detection system and is too difficult to achieve. The decision making technique incorporates two processes: training and testing. The training phase incorporates the ABC training module for generating profiles for all types of legitimate records and for storing these generated profiles in a database. In the testing phase, ABC detection module is used for testing the traffic. Figure 1, gives a brief description about the same.

3.4 Methodology

The proposed approach was tested in a simulated environment with the help of CloudSim [17]. In order to do so, first we characterized our attackers and we generated attackers and non-attackers cloudlets and labeled both of them. After this we extracted the features and data partitioning was done where some data were reserved for training and while the other for testing. Our approach has two phases for its implementation: training and testing. While in training phase we construct a normal profile using ABC algorithm, in testing phase main focus is on detecting the DoS attacks. In order to detect the attack, we have used centroids as classifiers. Centroids are generated for both attackers as well as non-attackers and distance is calculated with both to determine the attackers and non-attackers. At the end we evaluate the performance of the system. In order to do so, we calculate the mean or average accuracy of ABC. The results so obtained are then compared to QPSO. The flowchart in Fig. 2 gives a brief description about it.

Fig. 2 Flowchart



4 Experiments and Results

In our research work, main aim was to prove the efficiency of artificial bee colony optimization approach for the detection of denial-of-service attack in a cloud environment. The results obtained shows that ABC is efficient enough to do the same.

In our research we have compared the efficiency of ABC and QPSO. Figure 3 shows that ABC is successfully detecting maximum attacks with a rate of 75–80 %. The average accuracies were found for ABC and QPSO using (3). A total of 10 readings were taken.

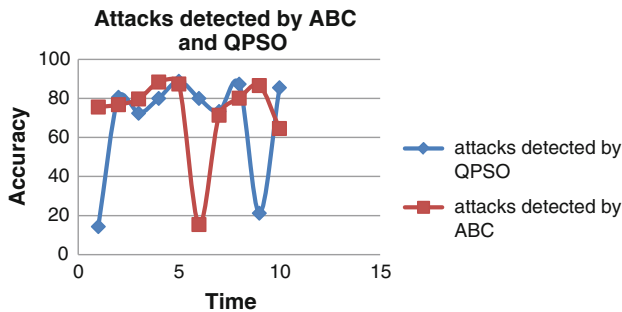


Fig. 3 Traffic detected by ABC and QPSO

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Total number of values}} \quad (4)$$

The average detection rate observed for ABC was 72.4 % while that for QPSO was 68.3 %.

Thus, from the above we can conclude that the average detection rate of ABC is far much higher than that of QPSO.

5 Conclusion and Future Work

Cloud computing provides a lot of advantages to its user to improve their conventional system. However, security should be alongside implemented to improve the performance and functionality. One of the serious threats come to cloud are from denial-of-service attack as this attack is easy to launch but difficult to stop. This research work has showed that artificial bee colony optimization, a swarm approach, is useful in detecting denial-of-service attack in a cloud environment. The proposed approach was carried out in a simulated environment using CloudSim [17]. The proposed approach also shows that it is able to detect most of the attacks in a very short period of time. This approach was further compared with quantum-inspired PSO and was found to be better. The results achieved for testing and training data sets were found to be 72.4 and 68.3 % for ABC and QPSO, respectively. In future, we will set up to work with real-world data and attacks to fine tune our system.

References

1. Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011).
2. Top Threats Working Group. "The notorious nine: cloud computing top threats in 2013." Cloud Security Alliance (2013).
3. Akbarabadi, Ahad, et al. "Client To Client Attacks Protection in Cloud Computing By a Secure Virtualization Model." *Journal of Next Generation Information Technology* 4.8 (2013).
4. Chen, Shuo, et al. "Side-channel leaks in web applications: A reality today, a challenge tomorrow." *Security and Privacy (SP)*, 2010 IEEE Symposium on. IEEE, 2010.
5. Joshi, Bansidhar, A. Santhana Vijayan, and Bineet Kumar Joshi. "Securing cloud computing environment against DDoS attacks." *Computer Communication and Informatics (ICCCI)*, 2012 International Conference on. IEEE, 2012.
6. Iftikhar A., Azween B. A., Abdullah S.A., (2009), "Application of Artificial neural Network in Detection of DoS attacks," *SIN'09*, Oct 6–10.
7. Mahajan, Ratul, et al. "Controlling high bandwidth aggregates in the network." *ACM SIGCOMM Computer Communication Review* 32.3 (2002): 62–73.
8. Lo, Chi-Chun, Chun-Chieh Huang, and Joy Ku. "A cooperative intrusion detection system framework for cloud computing networks." *Parallel processing workshops (ICPPW)*, 2010 39th international conference on. IEEE, 2010.
9. Lua, Ruiping, and Kin Choong Yow. "Mitigating ddos attacks with transparent and intelligent fast-flux swarm network." *Network*, IEEE 25.4 (2011): 28–33.

10. Anitha, E., and S. Malliga. "A packet marking approach to protect cloud environment against DDoS attacks." *Information Communication and Embedded Systems (ICICES)*, 2013 International Conference on. IEEE, 2013.
11. Reddy, Pallavali Radha Krishna, and Samia Bouzefrane. "Analysis and Detection of DoS Attacks in Cloud Computing by Using QSE Algorithm." *High Performance Computing and Communications*, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), 2014 IEEE Intl Conf on. IEEE, 2014.
12. Binitha, S., and S. Siva Sathya. "A survey of bio inspired optimization algorithms." *International Journal of Soft Computing and Engineering* 2.2 (2012): 137–151.
13. Karaboga, Dervis, and Celal Ozturk. "A novel clustering approach: Artificial Bee Colony (ABC) algorithm." *Applied soft computing* 11.1 (2011): 652–657.
14. Xiang X, Zhou W, Guo M., (2009), "Flexible deterministic packet marking: an IP Trace Back system to find the real source of attacks," *IEEE Transactions on Parallel and Distributed Systems*, Volume 20, No 4, pp 567–80.
15. S.J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, "Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project," *Proc. DARPA Information Survivability Conf. and Exposition (DISCEX '00)*, vol. 2, pp. 130–144, 2000.
16. D.E. Denning, "An Intrusion-Detection Model," *IEEE Trans. Software Eng.*, vol. TSE-13, no. 2, pp. 222–232, Feb. 1987.
17. <http://www.cloudbus.org/cloudsim/>.

Multi-cavity Photonic Crystal Waveguide-Based Ultra-Compact Pressure Sensor

Shivam Upadhyay, Vijay Laxmi Kalyani
and Chandrababha Charan

Abstract In this paper, we proposed an ultra-compact pressure sensor. It is designed using silicon photonic crystal waveguide with the multiple cavities. For better light confinement and simplicity in fabrication ‘air holes in slab type’ structure is used. For the propagation of light, transverse magnetic (TM) polarization mode is considered. The combination of silicon waveguide and multi-cavities gives high quality factor. The designed sensor is based on the principle of resonance wavelength. Applied external pressure changes the optical and electronic property of sensor thus resonance wavelength of sensor is shifted. It works in the conventional (c) band and short (s) band of communication system. The proposed design has very high quality factor of 1720 and sensitivity of 0.50 nm/GPa. All designing work is performed using layout designer tool and simulation work is performed using finite-difference-time-domain (FDTD) method and plane wave expansion (PWE) method.

Keywords Microelectromechanical system (MEMS) • Micro pressure sensor • Complementary metal–oxide–semiconductor (CMOS) • Finite-difference time-domain method (FDTD) • Plane wave expansion method (PWE) • Magnetic field distribution

1 Introduction

In present days, micro pressure sensor based on microelectromechanical system (MEMS) is widely used in various sensing applications, owing to their precise micro pressure measurement capability while keeping very compact size. To design

Shivam Upadhyay (✉) · V.L. Kalyani · Chandrababha Charan
Government Mahila Engineering College, Ajmer, India
e-mail: shivamupadhyay920@gmail.com

V.L. Kalyani
e-mail: kalyanivijaylaxmi@gmail.com

Chandrababha Charan
e-mail: charanchandrababha@gmail.com

ultra-compact (very small) and highly wavelength selective optoelectronic devices, photonic crystal is used as a material. Photonic crystal-based pressure sensor is dependent on the resonance wavelength principle. With the application of external force or pressure, its refractive index is changed which will change the resonance wavelength of sensor. Micro pressure sensor senses the pressure from very small pressure to GPa [1]. Photonic crystal (Phc) is the nanometer scale optical structure with the capability of confinement, controlling and manipulation of light. The very broad categories of application of photonic crystal are optical waveguide, microscopic optical cavities and photonic band gap (PBG) structure-based devices, etc. Photonic crystal-based optical technologies are also implemented in chemical and biochemical fields [1, 2]. In the sensing field, optical technology-based sensor are already implemented. All these are designed using directional coupler, Bragg grating or ring resonator and based on the principle of homogeneous sensing. Homogeneous sensing is related to refractive index modification and surface sensing is related with the thickness change of biomolecular layer which is immobilized on surface. The optical sensors designed for the detection and quantification of chemical, sensing of pressure, force, displacement, investigation of biochemicals and its interaction with the system of cellular are developed and it is still a field of extensive research [4, 5]. The optical sensor based on 2D-silicon (Phc) platform with group of air holes is more sensible because surface state of air holes modified the local electromagnetic fields of propagated wave. Thus, the sensor based on air holes type structure is very sensitive towards the small change in refractive index and implemented mostly for designing of physical, chemical and biochemical sensors. This type of structure is practically implemented for the fabrication process, due to its properties such as better light confinement capability in both vertical and lateral direction and easy in fabrication process. The complementary metal-oxide-semiconductor (CMOS) fabrication technology based on lithography and etching is used for making (Phc)-based devices using silicon on insulator wafer. For the fabrication of photonic crystals, diverse materials such as semiconductors, polymers, oxides and porous silicon are used [6, 7]. In this paper, we proposed a multi-cavity photonic crystal waveguide-based ultra-compact pressure sensor. This sensor can measure the GPa range of pressure and it has a very compact size, in the range of ultra. The designing part of sensor includes multiple cavities with the silicon waveguide.

2 Literature Review

Recently Olayee et al. have demonstrated high resolution pressure sensor based on the silicon rods suspended in air type structure with the principle of refractive index sensing. In this externally applied pressure, changes the effective refractive index of sensor. It detects the pressure from 0.1 to 10 GPa with the quality factor 1410 [8]. Leili et al. have designed a high sensitive double-hole defects refractive index sensor. Its layout configuration is consisting of the two waveguides coupled with

micro-cavities [9]. Stemo et al. demonstrate a force sensor based on photonic crystal silicon waveguide with the micro-cavity [10]. Lee et al. have proposed a novel nano mechanical sensors using silicon 2D photonic crystal, with the concept of resonance wavelength. In this resonance wavelength of output spectrum shows sensitivity toward the change in the dimension of air holes and defected length due to mechanical deformation [11]. Levy et al. demonstrates a displacement sensor based on the principal of planar photonic crystal waveguide (PHCWG) alignment. The light intensity of output is changed according to the variation in alignment accuracy. Suh et al. reported a displacement sensor based on the coupling of two photonic crystal slabs and form febray-perot cavity like structure [12]. Xu et al. has given a micro displacement sensor using photonic crystal with the codirectional coupler structure. The coupler has fixed and movable photonic crystal structure and it detects very small displacement between both the crystal structures [13].

3 Operating Principle

For the sensing mechanism, the effect of applied pressure on the electronic and optical properties of photonic crystal is considered. These properties are energy gap and effective refractive index of crystal. When any deformation or external pressure is applied on the sensor surface the complete pressure is distributed in the form of strain on the sensor surface thus sensor structure is compressed, which makes some changes in the electronic and optical properties of a sensor such as change in a refractive index and photonic band gap. This shifts the resonance wavelength of a device. In photonic crystal, the band gap property of crystal strongly depends on the refractive index, radius to lattice constant (r/a) ratio and lattice constant. When pressure is applied on it, the shape of air holes and refractive index of material is change. Thus at different pressures the normalized transmission spectrum of sensor is shifted. In this section, the effect of applied pressure on the refractive index of sensor material is calculated. For the calculation of refractive index of a sensor, when different pressures are applied, the optical tensor coefficient and optical tensor equation is use for the calculation purposes [14]:

$$\begin{bmatrix} n_{xx} \\ n_{yy} \\ n_{zz} \\ n_{yz} \\ n_{xz} \\ n_{xy} \end{bmatrix} = \begin{bmatrix} n_0 \\ n_0 \\ n_0 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} c_1 & c_2 & c_2 & 0 & 0 & 0 \\ c_2 & c_1 & c_2 & 0 & 0 & 0 \\ c_2 & c_2 & c_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{bmatrix}$$

Where refractive index along the ij direction is n_{ij} , n_0 represents refractive index of a sensor at 0 GPa pressure and σ_{ij} is the pressure along the ij direction. Now assume that the pressure is applied only in one direction, then

$$\sigma_{xy} = \sigma_{xz} = \sigma_{yz} = 0 \quad (1)$$

$$\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma \quad (2)$$

Thus applied pressure reduces the refractive index of sensor material. Then

$$n = \ddot{n}_0 - (c_1 + 2c_2)\sigma \quad (3)$$

where c_1 and c_2 are defined as:

$$c_1 = n_0(P_{11} - 2V \cdot P_{12})/(2E) \quad (4)$$

$$c_2 = n_0^3(P_{12} - V \cdot (P_{11} + P_{12}))/ (2E) \quad (5)$$

where E = Young's modulus constant, V = Poisson's ratio and P_{ij} = Strain optic constant.

4 Layout Configurations

A class of natural materials, in which refractive index of a material is periodically modulated is known as (Phc). By perturbing the internal structure of crystal, the quantum bundle of photons is propagated inside the photonic band gap of (Phc). Photonic crystal-based waveguide is a planar (Phc) with a line defect. Line defect is formed by removing the row of air holes. The main advantage of conventional photonic crystal wave guide is its light confinement capability, it is provided in lateral direction by photonic crystal and in vertical direction by total internal reflection (TIR). On the other side, to achieve very high quality factor some air holes, i.e. a point defect or nano cavity is also created in the structure and form a resonator. This high quality factor provides very sharp peak at resonance wavelength. In our proposed design, we use the same designing principle to achieve very high quality factor and implement this structure for the sensing of externally applied pressure. The fundamental design is based on the hexagonal lattice structure of silicon slab and group of surface air holes. The design has silicon photonic crystal waveguide embedded with some multi-cavity air holes in the structure. The lattice constant of structure is $a = 0.400 \mu\text{m}$, the radius of unit cell is $r = 0.345 \cdot a \mu\text{m}$, which is shown in Fig. 1. Here the waveguide is designed by removing row of air holes and cavities are formed by introducing point defects into the structure. The radius of cavities is $0.22 \mu\text{m}$. The complete structure has a cross-sectional area of $l \cdot w$, i.e. $8 \cdot 6.55 \mu\text{m}^2$. The Gaussian electromagnetic wave is generated using an optical source of wavelength $1.550 \mu\text{m}$ and output of sensor is detected using optical detector at the another end of waveguide inside the structure.

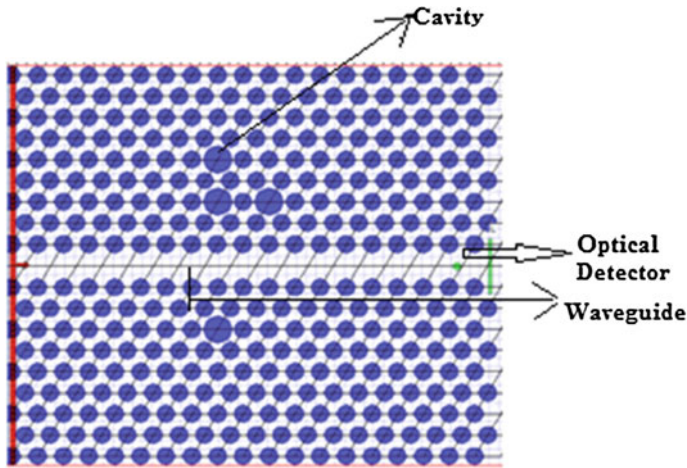


Fig. 1 Layout configuration of ultra-compact pressure sensor

5 Simulation Results

After layout designing of sensor structure, the band gap or light bandwidth of sensor is found using plane wave expansion (PWE) method with transverse magnetic (TM) polarization mode. To achieve good results, the perfectly matched layer (PML) boundary condition is considered in all four side of grid. The photonic band gap of designed layout is 0.55178–0.831137 eV and in terms of wavelength it is from 1.20 to 1.81 μm . The operating wavelength or resonance wavelength of sensor is 1.550 μm . All simulation work is performed using 2D-FDTD simulation method. It simulates the propagation of electromagnetic wave. The proposed pressure sensor structure measure the pressure from 0 to 4 GPa.

In Fig. 2, 2D magnetic field distribution of four cavities with the linear waveguide type structure is shown, when no pressure is applied on the sensor surface, i.e. at 0 GPa.

Figure 3 represent the effect of different pressure on sensor output from 1 to 4 GPa. The applied pressure distributes strain on the surface of sensor, which makes change into the effective refractive index of sensor thus the resonance wavelength of sensor is shifted.

Table 1 shows the complete performance of the proposed pressure sensor when pressure is applied on the sensor surface, the resonance wavelength of pressure sensor is shifted due to change in electronic and optical properties of sensor. Thus transmission power and wavelength shift is calculated.

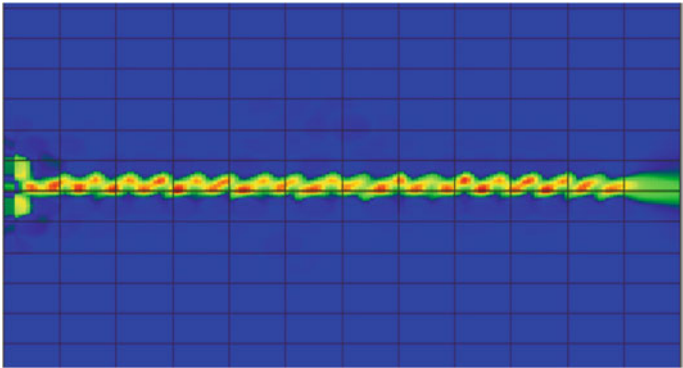


Fig. 2 2D Magnetic field distribution

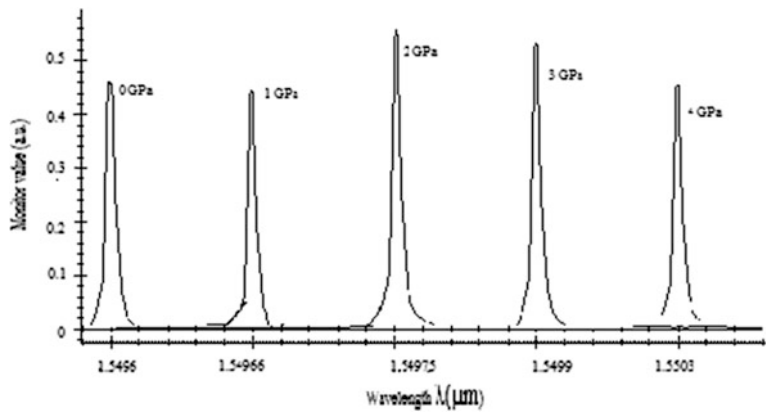


Fig. 3 Normalized transmission spectra of sensor at different pressures

Table 1 Performance analysis of pressure sensor

Applied pressure (GPa)	Effective refractive index	Resonance wavelength (μm)	Transmission power (%)	Wavelength shift ($\Delta\lambda$) (nm)
1	2.53985	1.54966	46.6	0.02
2	2.5797	1.54975	53.7	0.11
3	2.61955	1.54994	51.9	0.30
4	2.6594	1.55030	44.3	0.66

6 Analysis of Sensor

The proposed sensor is capable of measuring an externally applied pressure from 0 to 4 GPa with the better light confinement capability. This designed sensor can also be implemented for fabrication purposes because due to photonic slab type configuration as it has very less vertical leakage. The overall performance of sensor is analyzed by calculating the following parameters.

6.1 Quality Factor

Quality factor is the ratio of resonance wavelength (λ_0) to the full width at half maximum (FWHM) of resonator response. Its mathematical expression is given by

$$Q = \lambda_0 / \Delta\lambda$$

In the proposed design, the cavity structure has highest quality factor of 1720.

6.2 Sensitivity

The sensing capability of any sensor is given by the sensitivity. Sensitivity of a sensor is calculated by the ratio of shift in resonance wavelength to the change in refractive index. This is given by

$$S = \Delta\lambda / \Delta n$$

The proposed structure has the sensitivity of 0.50 nm/GPa.

7 Conclusion

In this paper, photonic crystal-based ultra-compact pressure sensor is proposed. The proposed pressure sensor is based on silicon waveguide with the cavities. The designed sensor has sensing and filtering capabilities. The cavities are providing filtration. Thus the proposed sensor provides accurate sensing output. Using hexagonal lattice structure with the lattice constant $a = 400$ nm and radius of cell is $r = 0.138$ μm . The refractive index of silicon slab is 3.5 or dielectric constant is 11.5. The proposed sensor follows the resonance wavelength sensing principle. In which due to external mechanical effects, the sensor structure is deformed either in terms of change in refractive index of material or change in shape of air holes, thus the resonance wavelength of sensor is shifted. At 0 GPa pressure the resonance

wavelength of sensor is 1.54964 μm and transmission power is 48 %. The quality factor of sensor is very high 1720 and sensitivity is 0.50 nm/GPa. This sensor works in conventional band (C-band) and short band (S-band) of communication system.

Acknowledgments The author would like to thank the referees for their constructive research work which helped to improve the quality of this paper. I also thank to Vijay Laxmi Kalyani, (Assistant Professor, Department of Electronics and communication, Govt. Mahila Engineering College, Ajmer) for her input on many aspects of this work.

References

1. Joannopoulos.: Photonic Crystals Molding the Flow of Light. In: Second Edition, Princeton University of press, 2008.
2. Yasuhide T.: Photonic Crystal Waveguide Based on 2-D Photonic Crystal with Absolute Photonic Band Gap. In: IEEE Photonics Technology Letters, vol. 18, no. 22, November 15 2014.
3. Fan X., Shopova S. I., White I. M., Suter J. D., Zhu H., Sun Y.: Sensitive Optical Biosensors for Unlabeled Targets: A review. In: Analytica Chimica Acta 620, 2008, pp. 8–26.
4. Liu Y., Salemink H. W. M.: Photonic crystal-based all-optical on-chip sensor. J. Optics Express, Vol. 20, No. 18, 2012, pp. 19912–19921.
5. Pacholski C.: Photonic Crystal Sensor based on Porous Sillicon: In: Journal of sensor, 2013 ISSN-1424-8220.
6. Monoik R., Streansky M., Sturdik E.: Biosensors - Classification Characterization and new Trends. In: Acta Chimica Slovaca, vol. 5, No. 1, 2012, pp. 109–120.
7. Ghoshal S., Mitra D., Roy S.: Biosensors and Biochips for Nonmedical Applications. In: Sensors & Transducers Journal, vol. 113, no. 2, 2010, pp. 1–17.
8. Olyae S., Asghar Dehghani A.: Nano-Pressure Sensor using High Quality Photonic Crystal Cavity Resonator. In: IEEE, IET 8th International symposium on communication Systems (CSNDSP 2012), Poznan University of Technology, Poznań, Poland, 18–20 July 2012.
9. Shiramin L.A., Kheradmand R., Abbasi A.: High-Sensitive Double-Hole Defect Refractive Index Sensor based on 2D- Photonic Crystal. In: IEEE Sensors Journal, vol. 13, no. 5, May 2013.
10. Stemeo T., Grande M., Passosea A., Salhi A., Vittorio M. D., Biallo D.: Fabrication of Force Sensor based on Two-Dimensional Photonic Crystal Technology. In: Microelectronic Engineering, vol. 84, no. 5–8, pp. 1450–1453, 2007.
11. Lee C., Thillaigovindan J., Radhakrishnan R.: Design and modeling of Nano-Mechanical Sensors using Sillicon 2D-Photonic Crystals. In: Journal of Lightwave Technology, vol. 26, no. 7, April 2008.
12. Levy O., Steinberg B.Z., Boag A., Nathan N.: Ultrasensitive displacement sensing using Photonic Crystal Waveguides. In Appl. Phys. Lett., vol. 98, pp. 033102, 2005.
13. Xu Z., Cao L., Gu C., Jin G., He Q.: Micro displacement sensor based on line-defect resonant cavity in Photonic-Crystal. In: Optics Express, vol. 14, no. 1, pp. 298–305, 2006.
14. Olyae S., Dehghani A. A.: High resolution and wide dynamic range pressure sensor based on 2D-photonic crystal. In: Photonic Sensors, vol. 2, 2012, pp. 92–96.

Role-Based Access Mechanism/Policy for Enterprise Data in Cloud

Deepshikha Sharma, Rohitash Kumar Banyal and Iti Sharma

Abstract Attribute-based encryption is the need of the hour due to rapidly growing shared data on cloud. The enterprises which are adopting cloud already have some access controls in place. Role-based access control is most popular of these. This paper proposes how RBSC₁ can be incorporated into an access mechanism/policy to be used with ABE. This enhances the motivation of enterprises towards putting their data on cloud. The mechanism is very efficient in terms of space and time. Also, it makes key revocation very easy.

Keywords ABE · RBAC · Access structure · Access control

1 Introduction

Increasing amount of digital information demands to be saved in large databases, which need to be secured at the same time. Encryption tools are the primary methods to ensure security of databases and the flow of information. A related problem is to manage the access of shared data. Many structures that help in control of data access have been adduced to resolve this purpose. Most of the time Role-based access control (RBAC) [1] is used as the access control model, which reduces the maintenance cost of classical access control. Implementing RBAC with encryption primitives was a challenge few years ago.

Attribute-based encryption (ABE) [2] successfully assimilates encryption and access control. In ABE, a well-defined attributes subset is used to generate IDs for user groups, and each ID corresponds to a secret key. The set of attributes might change from user to user, thus enabling an attribute-based access control. Access structures are used to map the attributes and access policies into keys. These structures are categorized into hierarchical [3], monotone and non-monotone access

Deepshikha Sharma · R.K. Banyal (✉) · Iti Sharma
Department of Computer Engineering, University College of Engineering, Rajasthan
Technical University, Kota, India
e-mail: rkbayal@gmail.com

structures [4, 5]. Generally, in ABE decryption is allowed only when attributes subset of the key matches with the attributes subset of ciphertext. The matching could be complete or partial, i.e. “k out of d” attributes of ciphertext match with private key. In key-policy ABE [6], policy tree is constructed from attributes associated with private keys rather than using lists of attributes along with their private keys. In ciphertext-policy ABE [4] access tree is constructed from attributes, and then private key is generated. Though, ABE is a solution, high cost of these policies and difficulty of key revocation limits their use.

RBAC models are popular when it comes to implementing access control over enterprise databases, but only the basic model has been used in combination with encryption [7].

This paper proposes a RBAC₁-based access mechanism with lower time and space costs. Moreover, it can be converted into an access policy with any ABE scheme. The aim is to find a solution to the issue of security of large amount of shared data residing on platforms like those provided through cloud. In such cases, the control mechanisms exist for the data, yet what kind of control structure will be used for encryption is an open question.

2 Literature Survey

The work related to the proposed work can be divided into two major parts—access control for attribute-based encryption with and access control used over databases, specifically RBAC and related models.

2.1 Access Structures Used in Attribute-Based Encryption

Access structures are used when data is shared by multiple parties, each having different kind of ownership/privilege. These are combined with cryptography schemes into access control policy. We have reviewed major techniques of access control in this section.

Fine-Grained Access Control Systems that allow fine-grained access control [6, 8] are flexible in dividing the access over data to individual user in the user set or group. These techniques employ a trusted server to store data. Control of access depends on software checks for authority, i.e. if the user is authorized for access or not. Data used in this scheme is classified according to the given hierarchy and data encryption takes place under the public key that is declared for the set of attributes.

Monotone Access Structure Monotone access structures are commonly used for encryption where large enterprises are divided into user’s sets or groups like role-based access control models [9]. Files in these models are arranged according to monotone B_f (Boolean expression) on attributes. Any user has the access of a file

f only if the attributes of that user satisfies the monotone B_f . The functions can use only positive “AND”, “OR” or “Thresholds (d out of k)”, and not “NOT”.

Bethencourt et al. [4], Cheung and Newport [10], Goyal [11], Balu et al. [12] have proposed such monotone access structures.

Non-Monotone Access Structure These access structures include negative attributes which use “NOT” as their Boolean symbol. In the year 2012, Nishanth and Devesh [5] gave an AND gate access structure which supports wildcard entries with negative attributes. This scheme relies on constant size key and ciphertext.

Hierarchical Structure Hierarchical ABE scheme provides features such as flexibility, scalability, fine-grained control of access by distributing data in the hierarchy. In HABE-trusted authority, public-key generator (PKG) has the responsibility to generate system parameters and then distribute them to the users in distributed system. PKG sends the master key with ciphertext and authorize other high level authorities as a hierarchy. Each consumer must have their own private key secret. HASBE scheme [3, 13, 14] accepts set-based attributes that are recursive in nature. These set-based attributes are used for data decryption. Key's depth is defined by number of recursions.

2.2 *Role-Based Structures for Access Control (RBAC)*

The fundamental idea of RBAC is to prevent the access of organization's important information from users. Because all the information of an organization is not useful for every user, so the users are assigned different roles and the access permission of data is given according to the roles. Roles and permissions are associated with each other. The concept of role and permissions is given by Sandhu [7], also called enterprise concept. So whenever RBAC is used, it is supposed to maintain the security according to an organization's perspective; thus it is divided according to the roles, permissions and responsibilities in that organization. Using model RBAC, database management becomes easier and secure.

Basic RBAC Model RBAC model (basic) uses role hierarchy concept thus named as senior-junior inheritance model [7, 9]. This concept of role hierarchy shows that any higher role in the hierarchy will automatically inherit all the rights and permissions to access of a role that is lower in the hierarchy, in real world a higher job position role in an organization has access to all rights to a lower job position role. Workflow is not considered in RBAC. In basic RBAC model 'task' is not separate from 'role'. Both concepts (role and task) imply in same manner. Different tasks in access control are done in same pattern even if having different characteristics. Before authorization is done, all the users are assigned the roles manually in RBAC. RBAC has four main (core) components that define all basic sets and the function that needs to be applied; are users (in the system), roles (assigned to users), objects (operations will be performed) and operations (performed on objects), respectively. These components have all the information that helps to take decisions during authorization. RB access control has two more

advanced properties (features), i.e. “Constraints” and “Role Hierarchy”. Role hierarchy feature provides flexibility to the system administrations in RB access control.

RBAC₀ The base model of RB access control is RBAC₀ that uses least requirements to work in an RBAC environment [9]. It is included in both advanced models RBAC₁ and RBAC₂. RBAC₀ have three basic properties/features:

- In the system, some predefined roles sets are present and unlike basic RBAC, partial order between roles is not present in RBAC₀. Any user has some roles set (allotted by admin) along with the set of permissions for object’s operation.
- User is allowed to develop some sessions, activate a set/subset of assigned roles in that session. The session is approximately same to a subject. Now user will become authorized owner of that session and only he can delete or change the session. But if the owner has deleted the session the activated role set will also be deleted.
- The owner of session also has the authority to active or deactivate any role in that session. Permissions in the session are only dependent upon the roles used within that session, only active roles and their permission can be determined.

RBAC₁ In RBAC₁ role hierarchy is present, i.e. permission can be inherited by some roles (not permitted) form other roles (permitted). RBAC₁ has all the features of RBAC₀. Only is that role sets are partially ordered in RBAC₁. All the permissions to the junior roles can be inherited by senior roles [15]. Also the owners of session are allowed to activate junior roles [16, 17]. A user can establish a session with any combination of roles junior to the user’s own role. Similarly, active session’s roles plus assigned junior roles have the permissions within that session.

RBAC₂ RBAC₂ is same as RBAC₀, the only difference is that in RBAC₂ requires some constraints to decide the acceptability for RBAC₀’s components [7, 9, 17]. Permission is given to only those values which have been accepted. When applying the RBAC₀ constraints using “user” functions, it will helpful in user assignment while using “role” functions will be helpful in assigning the permission. For acceptable values “acceptable” predicates will return else “not-acceptable” predicate will return as a constraint after applying the function.

RBAC₃ RBAC₁ and RBAC₂ are beyond comparison, so they do not match to each other [16–18]. All three models are included (RBAC₁ and RBAC₂ are joint as they are while RBAC₀ is add transitively) to form a new compound model RBAC₃. As RBAC₃ is formed with the combination of RBAC₁ and RBAC₂, it has both features, i.e. constraints and role hierarchy but this combination has several issues when it comes to implementation.

3 Proposed Structure

This section proposes a mechanism for access control which uses the properties of RBAC₁. This mechanism can be used with any identity or attribute-based encryption. We propose to call it “Role Based Control Mechanism/Policy (RBCP)”.

The assumptions of the RBCP are:

Assumption 1 Attributes have set-based values. The values are partially ordered.

Assumption 2 The maximum number of different access to be provided is known a priori.

Assumption 3 A role hierarchy exists as a constraint, implying inheritance property, i.e. all the senior roles automatically inherit the permissions from their junior roles. A permission assigned to a junior role must also be assigned to all senior roles.

This inheritance constraint eliminates some Boolean combinations of the attributes. Thus a n-ary tree of attributes with AND and OR operators gets pruned. Nodes which are irrelevant to the role hierarchy are deleted. This pruned tree now occupies less space. For example, if we have three attributes and two roles, a level in the access structure will represent a role and would have three branches for three attributes. This is illustrated in Fig. 1.

Total nodes in Fig. 1 are 12, giving a total of a possible access combination. Inheritance makes a few of these impossible. If the attributes have partial order, the user at level 1 with attribute 1, is junior most in the hierarchy. Thus, there is no need of checking its other attributes. This reduces the amount of work to be done to compare values of other attributes. Thus, a pruned tree is obtained as shown in Fig. 2.

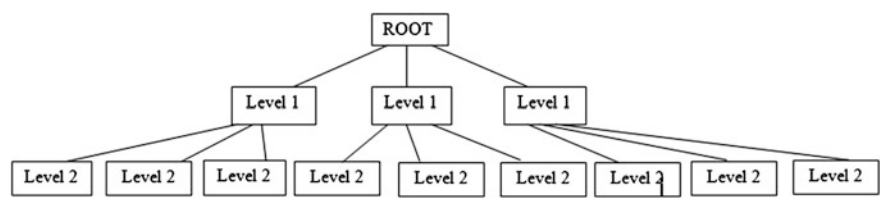


Fig. 1 Tree with three attributes

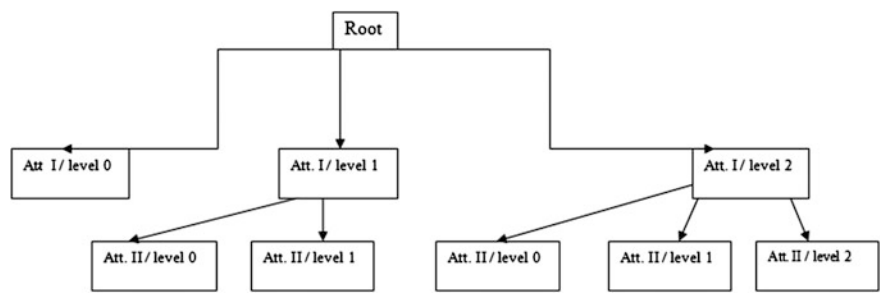


Fig. 2 Pruned tree

The proposed access structure constructs such pruned access tree, using only the minimum required combinations of attributes to generate an access.

The attributes derive values from a set of values which are partially ordered. When a user requests access, the strings of attributes needs to be compared which might require costly string comparison methods at implementation level. We propose to substitute these with integer operations. The partial order of the attribute values can be initialized to assign integer values to them. Moreover, computing the path of access in the structure each time a request arrives can be avoided if each path is assigned a unique value. Now all we have to do is find a function which can map integer values of the attributes to the unique access value. Let this function be

$$f(\text{attribute values vector}) = \text{access value}$$

If attributes exist at k different levels,

$$f(<a_{i1}, a_{i2}, \dots, a_{ik}>) = \text{acc_val}_i$$

for user i with attributes values all at level l .

If a function is implemented as a polynomial with fixed coefficient, its security is very low. Hence, we propose a function whose coefficients can be changed. For simplicity, we drop ' i ' which denotes user i ,

$$f(<a_1, a_2, \dots, a_k>) = c_1a_1 + c_2a_2 + \dots + c_ka_k$$

Here, all c_j 's are to be selected by order following conditions:

- (i) for all c_j , $c_j > k$
- (ii) for all c_j , $c_j + 1 = c_{j+1}$, $c_j \geq 1$
or
- (iii) for all c_j , $c_j = b_j + x$, where x is any integer ' $x \geq -1$ ' and ' b ' is any positive integer

Thus, coefficients are either successive positive integer or successive powers of a positive integer.

Changing coefficients is a decision which can be taken by the PKG, without the need of informing user about this decision. Thus, a key of user can be changed every session. Computation of access value is simply an integer calculation which is cost effective and are in changing of coefficients makes key revocation much easier. Table 1 shows how different coefficients produce altogether different values, but each attribute combination has a unique value, thus indicating a unique access path.

The values in the columns are computed using different coefficients as shown below

Table 1 Comparison between existing structure and proposed structure

Attribute value			Access value				
Level 1	Level 2	Level 3	F1	F2	F3	F4	F5
0	0	0	0	0	0	0	0
1	0	0	1	1	4	8	28
1	1	0	4	5	9	24	57
2	0	0	2	2	8	16	56
2	1	0	5	6	13	32	85
2	1	1	14	22	19	64	115
2	2	0	8	10	18	48	114
2	2	1	17	26	24	80	144
2	2	2	26	42	30	112	174
0	0	0	0	0	0	0	0

$$\begin{aligned} f_1 &= a_1 * 3^0 + a_2 * 3^1 + a_3 * 3^2 \\ f_2 &= a_1 * 4^0 + a_2 * 4^1 + a_3 * 4^2 \\ f_3 &= a_1 * 4 + a_2 * 5 + a_3 * 6 \\ f_4 &= a_1 * 2^3 + a_2 * 2^4 + a_3 * 2^5 \\ f_5 &= a_1 * 28 + a_2 * 29 + a_3 * 30 \end{aligned}$$

4 Analysis

From the point of view of implementation, the proposed access structure is only a mechanism, not a structure. So it does not occupy any memory, thus having $O(1)$ space complexity. The number of integers, k , involved in access mechanism affects the number of operation for key generation. Hence, time complexity is $O(k)$, which is constant for practical purposes.

Table 2 tabulates the points of comparison of proposal and other access mechanisms.

Table 2 Comparison between existing structure and proposed structure

	Existing access control structure	Proposed access control policy
1	Use Logical/Boolean techniques which are limited in expressive power	Implemented on RBAC model, inherits all the advantages
2	High space and time complexities	Very low Space and time Cost
3	Key revocation is a major issue in all the ABE schemes	Key revocation is possible by changing few coefficients
4	Same structure can be used for different groups but its key generation depends on the scheme	Same structure can be used for every group through unique key (set of coefficients)

5 Conclusion

The enterprises can be motivated towards moving their critical data on cloud, if we can ensure two things. First their model of fine-grained access matches the model of the organizational hierarchy. Second, security of data can be ensured even if it is shared among employees of the organizations.

In this paper, we adduce an access mechanism/policy to attain ABE that has built on “RB (Role Based) Access Control Model” for enterprise data. The mechanism has very low time and space requirement. It also gives an idea how easily key revocation can be implemented with integer computations. Moreover, the structure gives more flexibility, more variety for access control. It might not be used as a general structure, also might not be applicable to all the applications but for certain kind of role-based attributes, this mechanism is more suitable. It is most suitable for enterprise data where roles and hierarchies exist.

For future scope of this policy, we will be implementing this structure mechanism along with IBE (identity-based) and ABE (attribute-based) encryption schemes to resolve fine-grained control of access with better key revocation techniques.

References

1. Lan Zhou, Vijay Varadharajan, and Michael Hitchens: Secure Administration Of Cryptographic Role-Based Access Control For Large Scale Cloud Storage Systems. In: *Journal of Computer and System Sciences*, vol. 80, pp. 1518–1533 (2014).
2. V. Goyal, O. Pandey, A. Sahai, and B. Waters: Attribute-based encryption for fine-grained access control of encrypted data In: *Proceedings of the 13th ACM conference on Computer and communications security*, ACM, pp. 89–98 (2006).
3. Jeremy Horwitz and Ben Lynn: Toward hierarchical identity-based encryption. In: *Theory and Application of Cryptographic Techniques*, pp. 466–481 (2002).
4. J. Bethencourt, A. Sahai, and B. Waters: Ciphertext-Policy Attribute Based Encryption. In: *Security and Privacy IEEE Symposium on IEEE*, pp. 321–334 (2007).
5. Nishant Doshi, Devesh Jinwala: Constant Ciphertext Length in CP-ABE. In: *Advanced Computing, Networking and Security*, pp. 515–523 (2012).
6. V. Goyal, O. Pandey, A. Sahai, and B. Waters: Attribute Based Encryption for Fine-Grained Access Control of Encrypted Data. In: *ACM conference on Computer and Communications Security* (2006).
7. Ravi S. Sandhu: Role Based Access Control. In: *Advance in Computers*, vol. 46, pp. 237–286 (1998).
8. Junbeom Hur and Dong Kun Noh: Attribute-Based Access Control with Efficient Revocation in Data Outsourcing Systems. In: *IEEE Transactions On Parallel And Distributed Systems*, vol. 22, no. 7, pp. 1214–1221 (2011).
9. Ravi S. Sandhu: Role Based Access Control Model. In: *IEEE Computer*, pp. 38–47 (1996).
10. Ling Cheung and Calvin Newport: Provably Secure Ciphertext Policy ABE. In: *Proceedings of the 14th ACM conference on Computer and communications security*, pp. 456–465 (2007).

11. Vipul Goyal, Abhishek Jain, Omkant Pandey and Amit Sahai: Bounded Ciphertext Policy Attribute-Based Encryption. In: ICALP '08 Proceedings of the 35th international colloquium on Automata, Languages and Programming, Part II, pp. 579–591 (2008).
12. A. Balu and K. Kuppusamy: Ciphertext policy Attribute based Encryption with anonymous access policy. In: International journal of Peer to Peer Networks, vol. 1, no. 1, pp 1–8 (2010).
13. Craig Gentry and Alice Silverberg: Hierarchical id-based cryptography. In: Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security Springer-Verlag, pp. 548–566 (2002).
14. D. Boneh, X. Boyen, and E. Goh: Hierarchical identity based encryption with constant size ciphertext. In: Proceedings of Eurocrypt '05 (2005).
15. Sejong Oh and Seog Park: Task–role-based access control model. In: Information Systems, vol. 28, pp. 533–562 (2002).
16. Ravi Sandhu: Role Hierarchies and Constraints for Lattice Based Access Controls. In: Proc. Fourth European Symposium on Research in Computer Security, Rome, Italy (1996).
17. Elisa Bertino: RBAC Models-Concepts and Trends. In: Lecture Notes of Computer Science pp. 511–514 (2003).
18. Xin Jin: Attribute-Based Access Control Models And Implementation In Cloud Infrastructure As a Service. In: Phd Thesis of The University Of Texas At San Antonio (2014).

Big Data in Precision Agriculture Through ICT: Rainfall Prediction Using Neural Network Approach

M.R. Bendre, R.C. Thool and V.R. Thool

Abstract Weather forecasting with detailed and time-based information gathering is essential for future farming. This paper gives an abstract idea about big data in precision agriculture and how it discovers insights from big precision agriculture data through information and communication technology (ICT) resources for future farming. We proposed an e-Agriculture model for the use of ICT services in agricultural environment for collecting big data. Big data analytics provides a new insight to give advance decision support, improve yield productivity, and avoid unnecessary costs related to harvesting, use of pesticide, and fertilizers. The paper lists out the different sources of big data and types in precision agriculture, ICT-based e-Agriculture model, its future applications, and challenges. Finally, we have discussed rainfall prediction application using supervised and unsupervised method for data processing and forecasting.

Keywords Big data • Big data analytics • Precision agriculture • Information and communication technology

1 Introduction

In precision agriculture, historically generated data collected in structured and unstructured datasets lead to bigger size. The need of future farming is to improve the quality of agriculture products and services by reducing investment cost based on analysis of data. Big data can support wide range of precision agriculture functions for discovering intelligence and insights from data to address many new and important

M.R. Bendre (✉) · R.C. Thool · V.R. Thool
Shri Guru Gobind Singhji Institute of Engineering and Technology,
Vishnupuri, Nanded, India
e-mail: bendreminath@sggs.ac.in

R.C. Thool
e-mail: rcthool@sggs.ac.in

V.R. Thool
e-mail: vrthool@sggs.ac.in

farming decisions. In the agriculture sector, ICT plays an important role to provide new technologies for data generation, transformation, and management [1–3].

The researchers have an opportunity to discover knowledge from huge data. To discover relationship, find patterns and trends from the data for various management strategies. Thus, big data analytics applications in agriculture take advantage of the explosion in data to extract insights for making better decisions. The ICT provides information to farmers through mobile apps, SMS services, agriculture knowledge hubs, and new generation web applications. The ICT provides research equipments to the researchers for the precision agriculture, remote sensing such as GPS, GIS, devices, and data monitors.

This paper is organized in four sections. Section 2 provides background information, including types of data, characteristics of big data in precision agriculture, architectural model, management tools, and strategies. Section 3 provides case study and methodologies used for the prediction of rainfall using linear regression and neural network approach. Section 4 deals with results and discussion on big data in precision agriculture and rainfall prediction application. The last section brings main conclusions, and outlines possible directions for future work.

2 Big Data in Precision Agriculture

Maximum data in the agriculture sector are generated by the on-site farming, remote farming, or satellite farming called as precision agriculture.

Table 1 Types of data in PA

Sr. no.	Data type	Data sources
1	Historical data	Soil testing, crop patterns, field monitoring, yield monitoring, climate conditions, weather conditions, GIS data, and labor data
2	Agricultural equipment and sensor data	Remote sensing devices, GPS-based receivers, variable rate fertilizes, soil moisture, temperature sensor, farmers call records and equipment logs
3	Social and web data	Farmers and customers feedback, blogging sites like agro adviser, agriculture blogs, social media groups, web pages, and data from search engines
4	Publications	Farmers and customers feedback, blogging sites like agro adviser, agriculture blogs, social media groups, web pages, and data from search engines
5	Streamed data	Crop monitoring, mapping, drones, aircraft's, wireless sensors, smart phones, security surveillances
6	Business, industries and external data	Billing and scheduling systems, agriculture departments and other agriculture equipment manufacturing company

2.1 Types of Data and Characteristics

Agriculture big data are collected in the form of structured and unstructured from various homogeneous and heterogeneous sensing devices. Mainly, precision agriculture datasets have data related to crop patterns, crop rotations, weather parameters, environmental conditions, soil types, soil nutrients, geographic information system (GIS) data, global positioning system (GPS) data, farmer records, and agriculture machineries data, such as yield monitoring and variable rate fertilizers (VRT) [4, 5]. Typical types of PA datasets are given in Table 1.

2.2 Model of ICT for Precision Agriculture

Precision agriculture through ICT can be divided into different layers such as application layer, store and processing layer, and infrastructure layer. In the application layer, data acquisition tools, web-based solutions, and software's and development platforms are present. The storage and management of big data need a novel system and platform; today's cloud computing solutions provide such huge amount of storage and management [6, 7]. The distributed and parallel systems make a role in the data processing and management. The map-reduce based model can be used for the big data processing. Mainly ICT plays a role of data acquisition, management, and visualization over world wide in different applications [2]. The overall components in the precision agriculture are shown in the Fig. 1. Finally, infrastructure layer consisting of clustered network of sensors and systems are used to generate, access, and manage large amount of data.

2.3 Management Tools in Precision Agriculture

In the precision farming to gather data, process, visualize, and decision making the ICT plays an important role using technological tools such as hardware, software, and practices [8]. The various management tools and its applications are given in the Table 2.

2.4 Precision Agriculture Management Strategies

ICT approach in precision agriculture management gives novel technologies and platform to farmers, government departments, industries, and researchers. Following are the management strategies used in the precision agriculture [9–11].

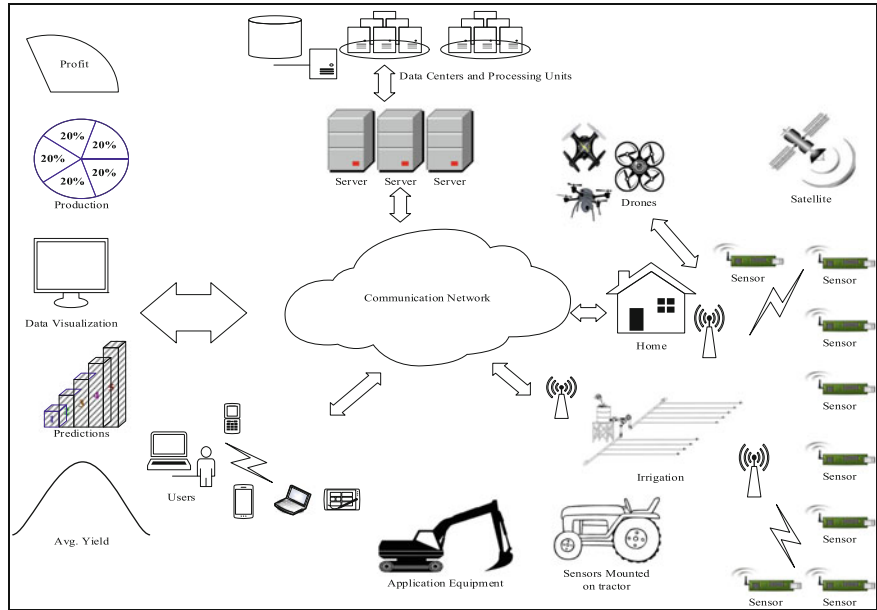


Fig. 1 Big data model of precision agriculture

Table 2 Management tools in PA

Sr. no.	PA management tools	Application
1	Global positioning system (GPS)	Position on the earth, real-time data acquisition, signal for measurement, increases the accuracy
2	Remote sensing (RS)	Collection of huge and variety of data, data acquisition, Sensing various parameters
3	Geographic information system (GIS)	Store yield maps, sensed data soil survey reports, soil nutrient levels, software and hardware modules for generation of the maps
4	Variable rate fertilizer (VRT)	Increase the soil fertility soil sampling is the recommended, ICT in agriculture to increase production

Farming Decision Support Big data analytics and ICT technologies help to acquire, understand, categorize, and discover information from large amount of data. Also predict future or recommend decisions to farmers and vendors at the point of precision agriculture.

Water Management Predictive data mining or analytic solutions over ICT can leverage water management and automatic irrigation system (e.g., as per soil humidity and new technology of irrigation) in real-time to improve best practices to crops.

Increase Productivity Web and mobile-based applications predict information from historical data, crop patterns, and weather data. Big data analytics and ICT

solutions can also support agriculture equipment companies and departments performing analysis over agricultural growth and productivity, to help and identify future farming trends.

Agriculture Disaster Management Big data analytics and ICT applications can support initiatives such as real-time management in precision agriculture, where it can mine knowledge from historical unstructured data, discover patterns to predict events that are harmful in farming. So, these decisions help in the disaster management in agriculture.

Policy, Financial and Administrative Analysis supports policy makers, service providers, companies, and government departments to decide future varieties, pesticides, and fertilizers.

2.5 Challenges

The main challenges are discovering knowledge and correlations from historical records, understanding big data, unstructured data in the right format, handling huge amount of statistical, imaginary and video data, handling data of crop monitoring through several sensors and their various interactions and communications, adoption and accessibility of new generation technologies for the individual farmer are expensive tasks [12]. In the agriculture community, lack of low technology knowledge needs more training, security, and management of ICT equipments [13].

3 Methodology and Case Study

3.1 Datasets

Weather forecasting for better agriculture decision and production in green zone or dry area is essential. In this case study daily min, max temperature, humidity, and rainfall data (Krishi Vidyapeeth Rahuri (KVR), Ahmednagar, India) of weather station from past 10 years were collected and analyzed. All the parameters used in the case study are rainfall in millimeter (mm), temperature in degree Celsius (°C), and humidity in percentage (%). KVR is 3 and 15 km far from two main rivers of the Ahmednagar district Pravara and Mula, respectively. This area of agriculture is in the green zone and half of the year, these rivers flow with water. So, the weather-based prediction and decision support study for the e-Agriculture in this area is important. The study data consist of daily weather parameters collected from KVR rain gauge between January 1, 2003 and December 31, 2013. Figure 2 shows the historical rainfall data used in the study. With the help of machine learning tools and techniques, it can predict future trends in the application [14].

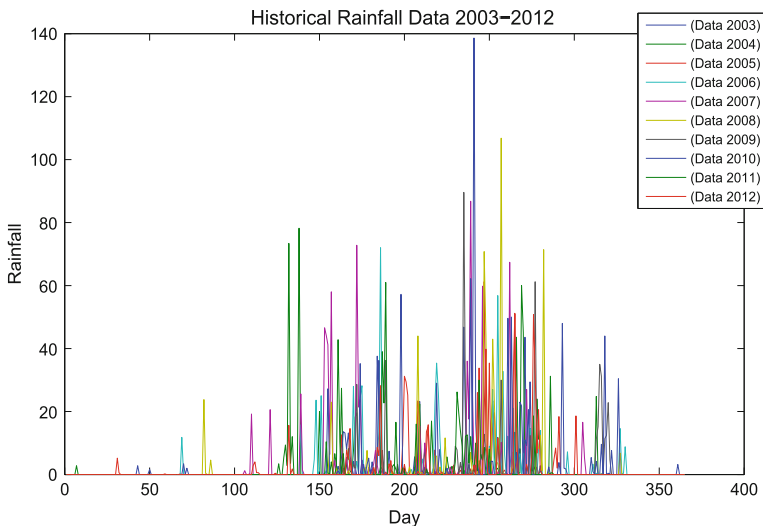


Fig. 2 Historical rainfall data

3.2 Regression Method Approach

Linear regression method is supervised learning for prediction of the future patterns from historical data items. The data collected from KVR station are stored in database by datastore functions and can access number of data vectors for processing. Where the vector x consisting of n numerical values $(x_1, x_2, x_3, \dots, x_{n-1}, x_n)$, and n is number of feature values of each data item in the dataset. The Eq. 1 shows general model used for the data processing.

$$\hat{y} = f(x) + \xi \quad (1)$$

In Eq. 1 the difference between actual and predicted value of target value is denoted by ξ . The predicted value for y is $f(x)$ and is indicated by $\hat{y}$ symbol. As linear regression could be used when there is a linear dependency between x and y . In this case, Eq. 2 shows the algorithm used to model y as a function of x .

$$\hat{y}_i = a_0 + a_1 x_i + \xi \quad (2)$$

To calculate the regression coefficients a_0 and a_1 the Eqs. 3 and 4 are used to minimize the error.

$$a_1 = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{y^2} - \bar{y}^2} \quad (3)$$

$$a_0 = \bar{y} - a_1 \bar{x} \quad (4)$$

Error term ξ is the difference between actual and predicted value of target variable. The objective is to minimize difference of actual and predicted values for all data items.

$$f(q) = \frac{1}{N} \sum_{i=1}^N \xi_i^2 \quad (5)$$

3.3 Neural Network Method Approach

Artificial neural network (ANN) [14] is mainly used for forecasting data in different problems. It uses activation functions to calculate threshold value for the different weights and bias. The common transfer functions in the ANN are tangent sigmoid pureline depicted. The formula of tansig and pureline transfer functions are expressed, respectively in Eq. 6.

$$F_k(s_k) = \frac{2}{1 + e^{-2s_k}} - 1 \quad (6)$$

Performance of ANN is classified into major groups based on the pattern of interconnection of neurons to propagate data. The values of weights and bias are responsible for the change in the performance. ANN should be configured to produce the desired output by adjusting the weights of interconnections among all neuron pairs. This process is called as training which is categorized into two main groups called supervised and unsupervised learning [14]. In supervised learning, ANN feeds with the learning patterns and adjusts the weights by comparison of the desired output with the actual output obtained from the input variables to achieve the minimum error.

4 Results and Discussion

This paper presents a model for big data and methodology for forecasting rainfall. The objective of this paper is to increase the accuracy of the forecasting using different weather parameters for the future precision agriculture. The proposed model can be used to gather big data using various ICT components. The methodology is illustrated using a case study for weather forecasting data collected between January 2003 and December 2013 at a weather station located in a green

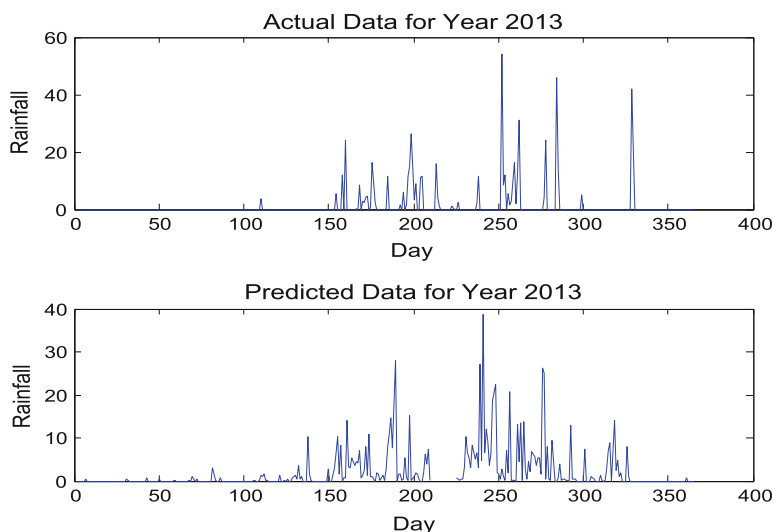


Fig. 3 Actual and predicted rainfall by linear regression

zone region of Ahmednagar. The historical data applied to processing model in this study, that is, rainfall through linear regression and neural network, minimize the processing time. The neural network algorithm processes fast and calculates better result as compared to the normal regression method. The model predicts rainfall and temperature values for the year 2013 and also compared actual and predicted values to minimize the error.

Neural network NARX has been applied in two training scenarios. One network training algorithm is the Levenberg–Marquardt optimization (`trainlm`) and the next network training algorithm is Bayesian regularization (`trainbr`). The algorithm functions were used for the neurons in the hidden layer and output layer, respectively. The weights and biases were adjusted based on the Levenberg–Marquardt and Bayesian regularization algorithms. The mean square error (MSE) and mean average error (MAE) were chosen as the statistical criteria for measuring the network performance. The algorithms are tested with different input values in percentage for training, testing, and validation. Figure 2 shows the input data of historical rainfall data used for the prediction study and different color denotes the year-wise data from January 1 to December 31. The best performance of the algorithm for given data is shown in the Figs. 3 and 4. Figures 5 and 6 show the regression output and error in the actual and predicted rainfall using the neural network approach.

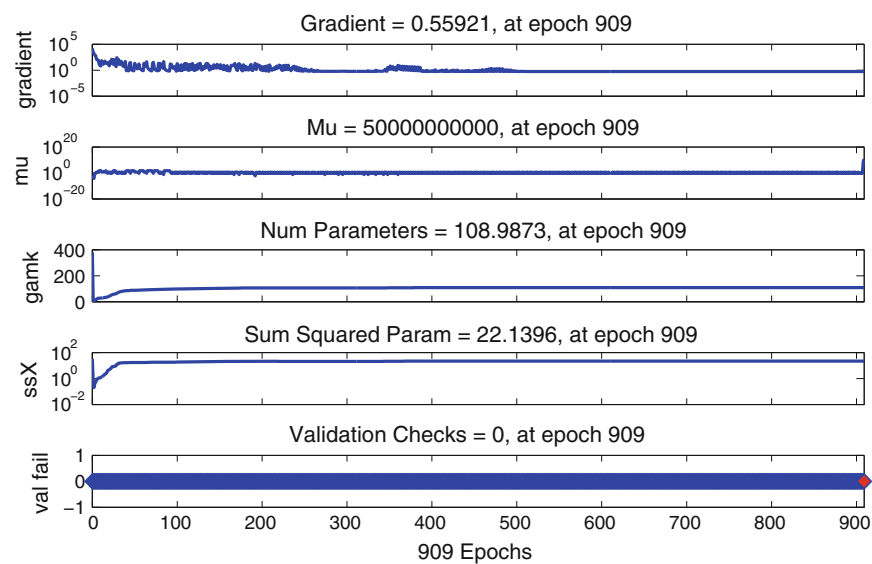


Fig. 4 Training state

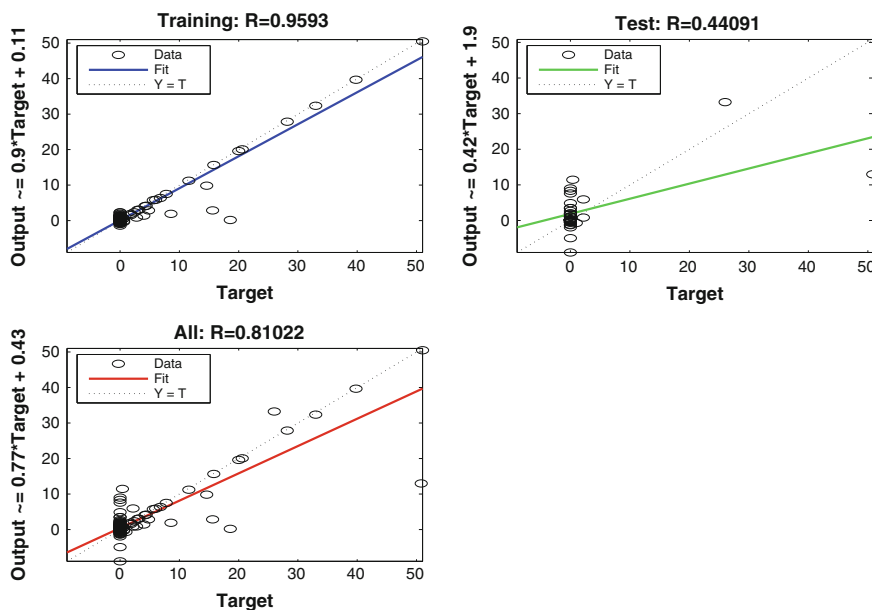


Fig. 5 Regression output

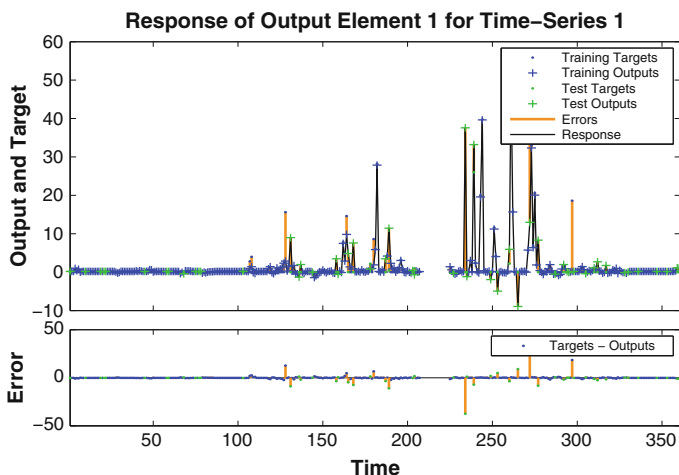


Fig. 6 Error, output, and target response

5 Conclusion

Big data analytics and ICT in agriculture are evolving technologies into a promising field for providing insight from very large data sets and improving productivity and reducing investment costs. Big data analytics and ICT have the potential to use novel technologies and platform to generate, collect, process, and visualize large data for future predictions and make decisions. In the precision agriculture, remote sensing devices play a vital role for data collection and real-time decision support. The results forecast using regression model and neural network model in this study show a considerable potential of data fusion in the field of crop and water management for applications such as precision agriculture. As per these results, the model predicts the rainfall in the region of case study. It suggests various decisions to farmers for deciding crop pattern and water management in future. It is solution for yield management and disaster management and to increase the gain of food production. In the future, we will see the rapid growth and use of big data analytics through ICT across the agriculture organization and the agriculture industry to increase yield production. Big data analytics and ICT applications in precision agriculture are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process to increase productivity of agriculture.

Acknowledgments The authors would like to thank KVR Rahuri for providing data and guidance for the study.

References

1. X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 1, pp. 97–107, 2014.
2. R. D. Ludena, A. Ahrary *et al.*, "Big data approach in an ict agriculture project," in *Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on*. IEEE, 2013, pp. 261–265.
3. R. D. Grisso, M. M. Alley, P. McClellan, D. E. Brann, and S. J. Donohue, "Precision farming. a comprehensive approach," 2009.
4. B. Brisco, R. Brown, T. Hirose, H. McNairn, and K. Staenz, "Precision agriculture and the role of remote sensing: a review," *Canadian Journal of Remote Sensing*, vol. 24, no. 3, pp. 315–327, 1998.
5. S. W. Searcy, *Precision farming: A new approach to crop management*. Texas Agricultural Extension Service, Texas A & M University System, 1997.
6. F. Awuor, K. Kimeli, K. Rabah, and D. Rambim, "Ict solution architecture for agriculture," in *IST-Africa Conference and Exhibition (IST-Africa), 2013*. IEEE, 2013, pp. 1–7.
7. D. Borthakur, "The hadoop distributed file system: Architecture and design," *Hadoop Project Website*, vol. 11, no. 2007, p. 21, 2007.
8. B. Venkatalakshmi and P. Devi, "Decision support system for precision agriculture," *International Journal of Research in Engineering and Technology*, vol. 3, no. 7, pp. 849–852, 2014.
9. R. Saravanan, *ICTs for Agricultural Extension: Global Experiments, Innovations and Experiences*. New India Publishing, 2010.
10. R. Khosla, "Precision agriculture: challenges and opportunities in a flat world," in *19th World Congress of Soil Science*, 2010.
11. A. McBratney, B. Whelan, T. Ancev, and J. Bouma, "Future directions of precision agriculture," *Precision Agriculture*, vol. 6, no. 1, pp. 7–23, 2005.
12. S. Nandurkar, V. Thool, and R. Thool, "Design and development of precision agriculture system using wireless sensor network," in *Automation, Control, Energy and Systems (ACES), 2014 First International Conference on*. IEEE, 2014, pp. 1–6.
13. K. Cukier and V. Mayer-Schoenberger, "Rise of big data: How it's changing the way we think about the world, the," *Foreign Aff.*, vol. 92, p. 28, 2013.
14. I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

Evaluating Interactivity with Respect to Distance and Orientation Variables of GDE Model

Anjana Sharma and Pawanesh Abrol

Abstract Eye gaze-based system requires the correct estimation and detection of gaze. Gaze-based input is processed to initiate different commands remotely in computing systems. However, various factors like interactivity, processor affinity, distance, orientations, light, image resolution, etc., affect the detection of gaze in the eye gaze models. In this research paper, work has been done to evaluate the processing time for the estimation of gaze direction for detecting the variations in interactivity for gaze-based models w.r.t. distance and orientation parameters of the subject. The experimental work has been done using gaze detection and estimation (GDE) model. The different results obtained by varying the number of processor affinities for finding the minimum CPU time taken by the model for different eye images. These results can further be used for improving and minimizing the interactivity time for enhancing the accuracy and performance of eye gaze-based systems.

Keywords Gaze-based models • Interactivity • Processor affinity • Orientation

1 Introduction

Interactivity is the ability of the system to generate response within the stipulated time frame. The faster the response the higher is the interactivity. Interactivity become a critical measure for the online working systems and networking systems for high-end interfaces or standalone offline systems. Processor interactivity is also an important aspect of central processor unit (CPU) utility and functioning. One of the important parameter for estimating the measure of the interactivity is the

Anjana Sharma (✉)

Department of Computer Science, GGM Science College, Jammu, J&K, India

e-mail: ltanjana@yahoo.com

Pawanesh Abrol

Department of Computer Science & IT, University of Jammu, Jammu, J&K, India

e-mail: pawanesh.abrol@gmail.com

computational time. This time may be the total time required by the CPU to complete the execution of the process. The interactivity depends on the execution time or the computation time taken by the program or the algorithm. This time is also called profile time. In large networks or cloud environment, the interactivity time lapse is more and may be a reason for failure or delay in loading of certain applications. The interactivity, dwell, and the profile time may play an important role in the applications of eye gaze-based systems. The minimum interaction time can enhance the productivity and the response time of the systems [1, 2]. This can be done by specifying the number of CPU or processors also known as processor affinity to specific processes thus enhancing interactivity. Once specified, a process will always be scheduled on the same processor thus ensuring that data structures required for it to operate are always available within that CPU's cache. One, two, or multiple CPUs can be assigned to execute different processes simultaneously. The process or thread will execute only on the designated CPU or multiple CPUs. It has been observed that the major dependency of the gaze-based systems is on the distance and orientation of the subject.

In this research paper, the experimental work has been done using gaze detection and estimation (GDE) model to evaluate the performance and interactivity time for gaze-based models with respect to distance and orientation parameters of the subject. The GDE model has been proposed using edge detectors and other morphological functions to find out the glint coordinates of different eye images of the subjects based on the coordinates of the glint detection [3]. The different results are obtained using varying the number of processor affinities of the CPU's. The processor affinity of single or multiple number of CPU's have been analyzed for finding the execution time taken by the model for different eye images. The execution of the algorithm for the single or multi CPU's time has been studied for single as well as multiple processors for finding the gaze direction and estimation. Certain applications like profiler in MATLAB programming environment are being developed to reduce the execution time taken by an application by debugging and optimizing code files by tracking their execution time. The profile records information about execution time, number of calls, parent functions, child functions, etc., and helps in debugging and reducing the code lines which are taking maximum execution time. The results obtained by GDE model can further be used for improving and minimizing the interactivity time taken for enhancing the accuracy and performance of different eye gaze-based systems.

The literature review is discussed in Sect. 2. The methodological approach to estimate the eye gaze direction is presented in Sect. 3. Experimental results are given on different orientation and distances for subjects in Sect. 4. The conclusion and further research directions are discussed in Sect. 5.

2 Literature Review

Some of the significant algorithms and models using the processor or CPU time in real-time applications are given by different researchers as presented below. These can also be used for the betterment of eye gaze systems.

The GDE model has been proposed by Sharma et al. using edge detectors and other morphological functions to find out the glint coordinates. However, the algorithm needs to be further studied for improved efficiency in terms of execution time at different distances and orientation of the subjects [3]. In the paper by Gidlof et al. the authors compares the dwell time, number of dwells and the total number of options attended by the participants in the search of a particular product of their choice amongst different alternatives. These times taken while selecting the products are the measures of information acquired from each product specifically [4]. The effectiveness of the study by the authors in a 3D virtual reality (VR) system is implemented by a VR modeling language. To determine the eye gaze position the authors perform objective and subjective tests. The average time is calculated for the predetermined positions in comparison with the conventional mouse with a keyboard. The amount of time elapsed is smaller using mouse at the initial stages, but is reduced by the proposed method as the number of trials increases [5]. Dynamic voltage scaling (DVS) and multiple non-DVS system devices has been adopted by the authors Yang et al. to reduce the energy consumption of the processor by slowing down the processor speed. The authors proposed energy efficient scheduling for periodic hard real-time tasks in a system to minimize the system energy consumption of a given set of real-time tasks executing in its worst case. The proposed algorithm can reduce the energy consumption both in the CPU and system devices [6]. While designing time-critical applications, schedulability analysis is used to define the feasibility regions of tasks with deadlines to find the best design within the timing constraints as observed by Zeng et al. The formulation of the feasibility region is based on the response time calculation. Approximation techniques have been used to define a convex subset of the feasibility region. These techniques are used in conjunction with a branch and bound approach to compute suboptimal solutions for optimal task period selection, priority assignment, or placement of tasks onto CPUs. The authors provide an improved and simpler real-time schedulability test that allows an exact and efficient definition of the feasibility region [7].

It is evident from the literature review that the speed is very crucial and the CPU time or interactivity can be studied by analyzing the execution of the processes by varying processor affinity. The processor affinity of one, two, or multiple CPU's and the profile time are significant variables which may affect the processing or CPU time of the running process.

3 Methodology

Based on the literature review, the present experimental study have been undertaken to analyze the impact of distance and orientation of the user on the interactivity of the GDE model in relation to the average CPU time taken. The objective of this research work has been done to evaluate the average processing time for the identification of gaze direction using different subjects. This evaluation is further analyzed for estimating the variations in interactivity for gaze-based models with respect to distance and orientation parameters of the subject. The complete work flow is given in Fig. 1a. The image database $DB[i]$ is created using a high-resolution camera C_D . For the present study, SONY NEX-5 ultra compact with 15 fps, 4592×3056 resolution, and 14.2 megapixels sensor fitted with large articulated 7.5 cm monitor have been used. The inputs images of six different subjects in an indoor environment in a laboratory under normal lighting conditions have been taken. The input image I_i is normalized by reduction in noise and further cropped to a size of 220×120 pixels for one eye for the uniformity of results producing I_b in bmp format [8]. Further, I_b is processed for removing unwanted regions or boundaries for the location of exact glint coordinates for the detection of gaze quadrant. The different CPU times are calculated and further analyzed subject wise, distance wise and orientation wise. An experimental set up has been created for the analysis of the GDE as shown in Fig. 1b. Each subject S looks at C_D from three different distances D of 4, 6, and 8 ft. The S maintains head stationary relative to the gaze camera while gazing at each of

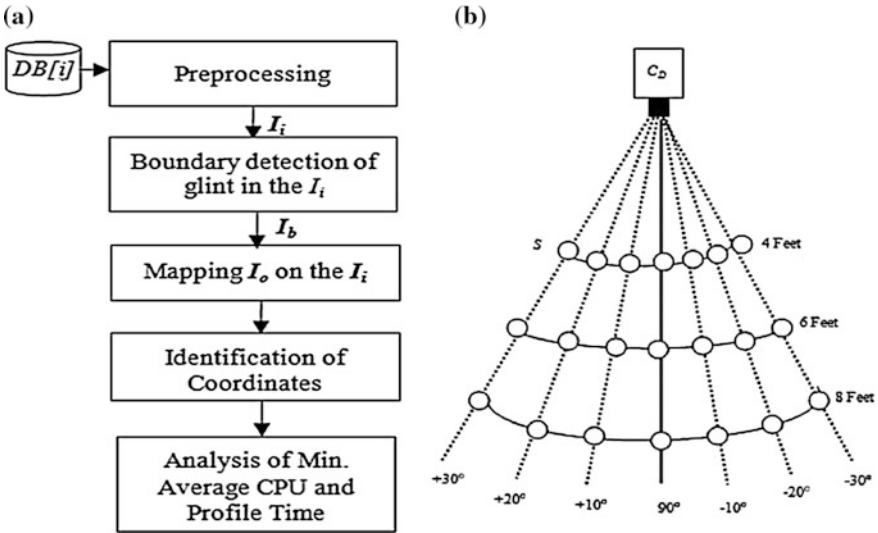


Fig. 1 a Work flow for evaluating interactivity with respect to distance and orientation. b Experimental setup for interactivity of GDE model

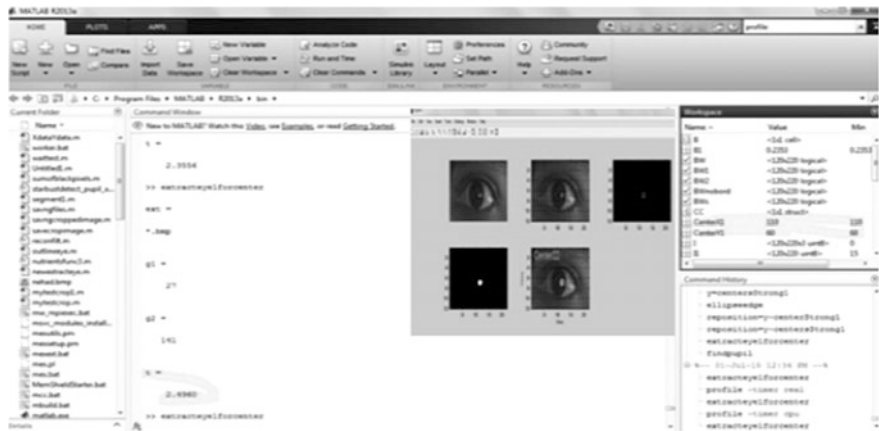


Fig. 2 Screen shot of MATLAB workspace

the distance for each of the seven different orientations. At each distance D , the orientation of S is changed with $\pm 10^\circ$ to obtain seven different locations.

The different gaze directions are detected at 21 different locations for each S . The experimental study has been done using more than 126 inputs in order to study the impact of the varying distance and orientation on the gaze detection model. The average CPU time with different processor affinities (one, two, and four) has been computed at different distances d and orientations θ . Experimental implementation of the used algorithm is done using MATLAB R2013 ver. 8.1.0.604 environment using a Windows7 64-bit Operating system, Intel® core i5 CPU, 2.40 GHz, 3 GB RAM with the Picasa version 3.9.137 Photo Viewer for editing images as per the requirement. Depending on the position of the coordinates of the glint, the model maps the gaze to the respective center quadrant for gaze detection as shown in Fig. 2. The average CPU time taken by all the input images with different processor affinity is denoted by T_C like with one CPU T_{C1} , for two CPU T_{C2} or for multiple CPU T_{C4} . The profile time for one and two CPU is denoted by T_{CP1} and T_{CP2} , respectively, as shown in Fig. 3. These different CPU times have been analyzed at each orientation θ and distance D .

4 Results and Discussion

As mentioned above, the processor interactivity is an important aspect of CPU utility and functioning. In this paper, effect of interactivity of eye gaze-based images in relation to the execution time has been analyzed for center gaze direction based on different distances and orientations as already mentioned above. The different results have been obtained from the experimental setup. The subjects are analyzed for minimum interactivity time of single CPU, two CPU, and Multi

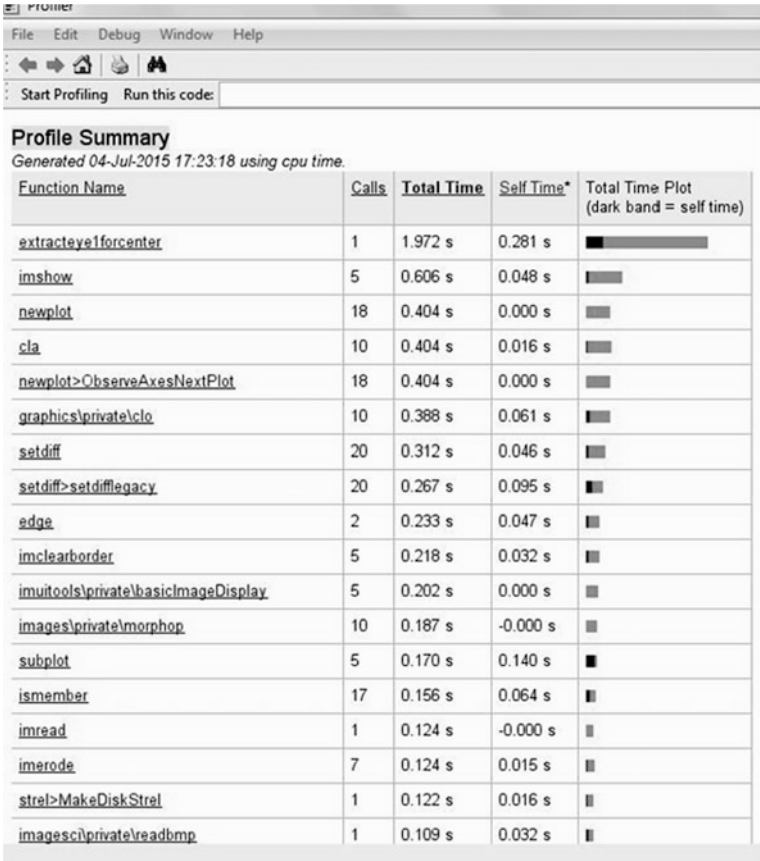


Fig. 3 Screen shot of MATLAB for profile time using CPU time

CPU's. Table 1 shows the average CPU time (T_{C1}) for one selected subject at different orientations and distances. The second column lists the execution time taken at $D = 4$.

The subsequent columns show the execution time at $D = 6$ and 8 ft distances at each orientation, respectively. The result (in bold) indicates the minimum average CPU time for all the orientations at 80° with a distance $D = 4$ and 6. At $D = 8$ the minimum average CPU time is at 120° . The results for the average CPU time for T_{C1} and T_{C4} for seven different orientations are listed in Table 2 taken by the six subjects. The second column and fourth column lists the execution time taken at each of the three distances at each orientation of all the six subjects. The subsequent third and fifth columns show the average of the execution time using T_{C1} and T_{C4} for the six subjects.

As observed from the table out of all the CPU times, the minimum average CPU time is at 80° (in bold) which marginally vary from the normal. The minimum average time for T_{C1} is 1.8599 in seconds and for T_{C4} is 2.2065 s. Similar results

Table 1 Single processor time T_{C1} at different distance and orientation for one subject

Degree	Distance			T_{C1} (avg.)
	4	6	8	
60°	1.8564	1.8564	1.9188	1.8772
70°	1.7784	1.8096	1.8252	1.8044
80°	1.7628	1.7628	1.8252	1.7836
90°	1.7940	1.7940	1.9032	1.8304
100°	1.8720	1.9032	1.8564	1.8772
110°	1.7784	1.8252	1.9032	1.8356
120°	1.7940	1.8096	1.7784	1.7940
TOTAL	12.6360	12.7608	13.0104	–
AVG	1.8051	1.8230	1.8586	–

Table 2 Average processor time T_{C1} and T_{C4} versus orientation and distance

Degree	T_{C1}	T_{C1} (avg.)	T_{C4}	T_{C4} (avg.)
60°	11.2788	1.8798	13.6240	2.2707
70°	11.2944	1.8824	13.4992	2.2499
80°	11.1592	1.8599	13.2392	2.2065
90°	11.3722	1.8954	13.8684	2.3114
100°	11.3415	1.8902	13.3952	2.2325
110°	11.4980	1.9163	13.6500	2.2750
120°	11.3880	1.8980	13.4836	2.2473

have been obtained for T_{C2} at $D = 6$ with the minimum average time 1.9344 s. The observation show the increase in execution time as the number of CPUs is increasing as the operating system will decide the assignment of the applications to which CPU for processing.

The average CPU timings of the single CPU and multi CPUs for the three distances have been obtained using six subjects as shown in Table 3. The second column and fourth column displays the total CPU time taken for the distances by the GDE model for all the six subjects. The third and fifth column is the average CPU time for all the six subjects at the respective distances. For both T_{C1} and T_{C4} , the minimum time is at $D = 4$ (in bold).

The minimum average time for T_{C1} is 1.8701 s and for T_{C4} is 2.2438 s. This indicates the increase in execution time as the number of CPUs is increasing. Result has also been analyzed for T_{C2} also and the minimum T_{C2} is at $D = 6$ is 1.9344 s. Besides the average CPU time for T_{C1} , the results for the profile time for all subjects have been calculated for single CPU (T_{CP1}) for the same subject have also been generated with the minimum CPU profile time T_{CP1} is at 80° as shown in Table 4.

Table 3 Average CPU time T_{C1} and T_{C4} versus distance

Distance	T_{C1}	T_{C1} (avg.)	T_{C4}	T_{C4} (avg.)
4	11.2209	1.8701	13.4628	2.2438
6	11.4442	1.9074	13.5809	2.2635
8	11.3344	1.8891	13.5675	2.2613

Table 4 Single CPU profile time T_{CP1} versus orientation

Degree	T_{CP1}	AVG
60°	11.8743	1.9790
70°	12.1207	2.0201
80°	11.8477	1.9746
90°	11.9770	1.9962
100°	12.1153	2.0192
110°	12.3277	2.0546
120°	12.1737	2.0289

Table 5 Two CPU profile time T_{CP2} versus distance

Distance	T_{CP2}	AVG
4	11.7399	1.9566
6	11.4111	1.9019
8	11.5364	1.9227

The aggregate profile time for all the subjects has been calculated to find the average profile CPU time T_{CP1} at each orientation θ . The first column is having seven degrees starting from 60, the second column displaying the total profile time by the CPU for the execution of the GDE model for all the six subjects. The last column is the average CPU times of all the six subjects at all the orientations. The minimum results at the 80° orientation marginally differ from the normal. The minimum average time for T_{CP1} is 1.9746 s. It can be observed that T_{CP2} have also generated the minimum average time in 1.9076 s. The minimum average CPU time results in a decrease in execution time as the number of CPUs increases.

Table 5 is showing the average of the profile time taken by the CPU for the three distances for all the six subjects. The last column is the average CPU profile time of all the six subjects. For T_{CP2} , the minimum average time is 1.9019 s at $D = 6$ quite similar as in the case of T_{C1} and T_{C2} . The results for T_{CP1} have also been prepared with the minimum average time for $T_{CP1} = 1.7255$ s and the minimum distance is at $D = 4$ showing the increase in the execution time as the number of CPUs increases.

All the average CPUs timings along with the profile timings for all the subjects at different orientations have been combined for the evaluation of the results as shown in Table 6.

Table 6 Average of different CPUs time versus orientation for six subjects

Average CPU time					
Degree	T_{C1}	T_{C2}	T_{C4}	T_{CP1}	T_{CP2}
60°	1.8798	2.0003	2.2707	1.9790	1.9207
70°	1.8824	1.9821	2.2499	2.0201	1.9159
80°	1.8599	1.9344	2.2065	1.9746	1.9076
90°	1.8954	1.9578	2.3114	1.9962	1.9337
100°	1.8902	1.9922	2.2325	2.0192	1.9419
110°	1.8902	1.9647	2.2750	2.0546	1.9623
120°	1.8980	1.9829	2.2473	2.0289	1.9074

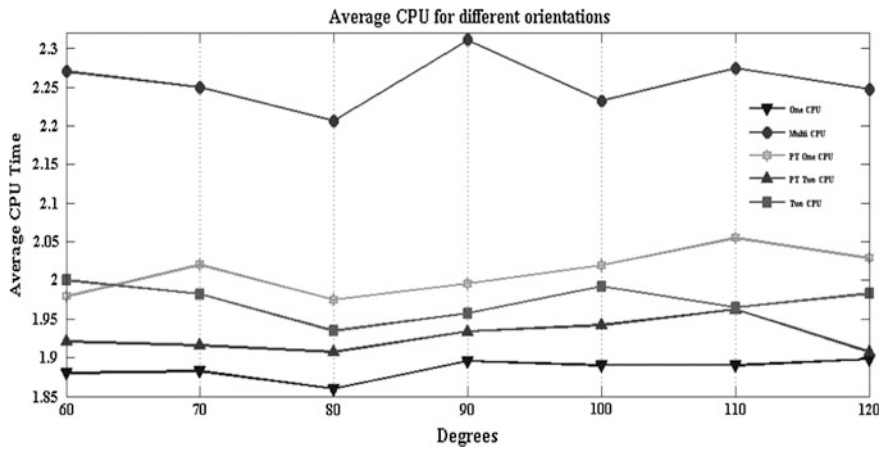


Fig. 4 Representation of average CPU times versus orientations for six subjects

The first column in the table is for different degrees starting from 60 up to 120°. The next three consecutive columns represent average CPU time with different processor affinity T_{C1} , T_{C2} and T_{C4} . The last two columns depict the profile time T_{CP1} for one CPU and T_{CP2} for two CPU taken by the GDE model for finding the gaze direction. The range for the CPU time starts from 1.8798 to 2.3114 s. The final results are also in compliance with the individual results generating the minimum average CPU time at 80°. The results show the minimum CPU time is at 80° which vary marginally from the normal. The results have been graphically represented in Fig. 4 showing all the minimum average CPU time at 80°.

Table 7 shows the average CPU time at different distances for six subjects. All the columns in the table representing the average CPU time with different processor affinity except the first one representing the three distances $D = 4, 6$ and 8 . The last two columns show the processor affinity with profile time. It has been observed from the analysis that out of the different five CPU timings the minimum average time is at $D = 4$ for three CPU's T_{C1} , T_{C4} and T_{CP1} . The remaining CPU's (T_{C2} and T_{CP2}) show minimum time at $D = 6$. However, the difference is very minor between the average time of 4 and 6 ft. The average CPU time ranges from 1.7255 to 2.2635. The average CPU time verses subject distance from the C_D has been graphically shown in Fig. 5. The graphical representation also showing the minimum time for three CPU T_{C1} , T_{C4} , and T_{CP1} at $D = 4$ and others at $D = 6$.

Table 7 Average of different CPUs time verses distance for six subjects

Average CPU time					
Distance	T_{C1}	T_{C2}	T_{C4}	T_{CP1}	T_{CP2}
4	1.8701	2.0091	2.2438	1.7255	1.9535
6	1.9074	1.9277	2.2635	2.1185	1.9080
8	1.8891	1.9837	2.2613	2.1908	2.2613

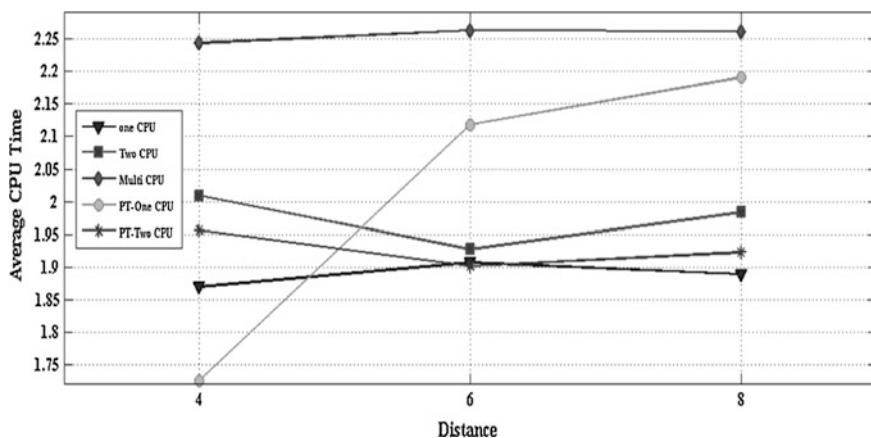


Fig. 5 Representation of average CPU times versus distance for six subjects

The analysis of the results shows that the interactivity of the CPU plays an important role in the direction and estimation of gaze.

The analysis indicates that the 80° orientation in all the cases is displaying the minimum average CPU time. Better accuracy for gaze-based input may be obtained at a maximum of 10° variation in the orientation with respect to normal at a distance $D = 4$ for T_{C1} , T_{C4} , and T_{CP1} . However, remaining two distances of 6 and 8 ft also generate acceptable results.

5 Conclusion

The gaze direction estimation model GDE is used for finding the minimum interactivity time required for its execution in order to minimize its complexity. The interactivity is taken on the basis of number of processors and the profile time using CPU time in MATLAB environment. The inputs have been taken with six different subjects in an indoor environment in a laboratory under normal lighting conditions. The input image of the subject is taken at seven different orientations with three distances of 4, 6 and 8 ft in order to study the effects of distance and orientation on the execution time of the GDE model. The best of the two eyes image with proper glint is taken for the analysis of the result. The interactivity results for all the seven orientations from 60° to 120° at three different distances are observed. More than 126 inputs have been generated for studying the impact of the distance and orientation on the interactivity of the gaze detection model.

The analysis indicates that the 80° orientation in all the cases is displaying the minimum average CPU time. Other results indicate that at the distance of 4 ft the average CPU time for T_{C1} , T_{C4} , and T_{CP1} is less whereas for T_{C2} and T_{CP2} the average time appears to be less at 6 ft. Better accuracy for gaze-based input may be

obtained at a maximum of 10° variation in the orientation with respect to normal and at a distance of 4 ft. However, distance of 4–6 ft also generates acceptable results. The results indicate an increase in execution time as the number of CPUs is increasing. The study may be conducted with more number of subjects at different distances and orientations in order to enhance the working range efficiency of gaze based systems.

References

1. Hansen, D. W. and Ji, Q.: In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. vol. 32, 3, pp. 478–500, In: IEEE Computer Society (2010).
2. Sharma, A. and Abrol, P.: Eye Gaze Techniques for Human Computer Interaction: A Research Survey. *Int. J. of Comput. Applicat. (IJCA)*, 71, pp. 18–29 (2013).
3. Sharma, A. and Abrol, P.: Comparative Analysis of Edge Detection Operators for Better Glint Detection. In: *IEEE Proc. of IX INDIACOM, 2nd Int'l Conf. on Computing for Sustainable Global Development*, pp. 973–977 (2015).
4. Gidlöf, K., Wallin, A., Dewhurst, R. and Holmqvist, K.: Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. *J. Eye Movement Research* 6(1):3, 1–14 (2013).
5. Choo, C., Lee, J. W., Shin, K.Y., Lee, E. C. K., Park, R., Lee, H. and Cha, J.: Gaze Detection by Wearable Eye-Tracking and NIR LED-Based Head-Tracking Device Based on SVR. *J. ETRI*, vol. 34, 4, pp. 542–552 (2012).
6. Yang, C., Chae, J., Hung, C. and Kuo, T.: System-Level Energy-Efficiency for Real-Time Tasks. In: *10th IEEE Intl. Sym. On Object and Component-Oriented Real-Time Distributed Computing* (2007).
7. Haibo, Z., Di Natale, M.: An Efficient Formulation of the Real-Time Feasibility Region for Design Optimization. In: *IEEE Computer transactions*, vol. 62 (4), pp. 644–661 (2012).
8. Sharma, D. and Abrol, P.: SVD Based Noise Removal Technique: An Experimental Analysis. *Int. J. of Advanced Research in Computer Science*, 3(5), pp. – 214–218 (2012).

Comparative Evaluation of SVD-TR Model for Eye Gaze-Based Systems

Deepika Sharma and Pawanesh Abrol

Abstract Eye gaze techniques require a number of eye inputs which can be taken by a capturing device like digital camera, webcam, etc. These eye inputs are usually in the form of digital images. With some powerful software, features can be removed or replaced in a digital image without any detectable trace and such operations are called tampering. Tampering also includes the addition of noise which is an unwanted data applied to image to disturb its basic features and results in false information. Therefore, it becomes essential to identify the tampering extent for such images. In this research, SVD-based noise detection and removal (SVD-TR) model has been applied to remove noise (salt-pepper and Gaussian) from eye gaze-based image database. The results show that SVD-TR model removes noise effectively from eye gaze-based image database. To check the efficiency of SVD-TR model, the results obtained are compared with median filter.

Keywords Eye gaze-based systems · Tampering · Gaze estimation · SVD · Noise · Salt-pepper · Gaussian noise

1 Introduction

Eye gaze is the process of measuring either the point of gaze or the motion of an eye relative to the head. The gaze point is estimated after acquiring the movement of the eye [1]. It requires significant levels of accuracy and estimation so that certain desired instructions can be executed by the computing system. Eye gaze techniques require a number of eye inputs, which can be taken by a capturing device like a digital camera, a webcam, single or multiple cameras, etc. The eye inputs may be the contour of the visible eyeball region, intensity distribution of the pupil(s), iris

Deepika Sharma (✉) · Pawanesh Abrol

Department of Computer Science & IT, University of Jammu, Jammu, J&K, India
e-mail: deepika.sharma231985@gmail.com

Pawanesh Abrol

e-mail: pawanesh.abrol@gmail.com

and cornea, as well as their shapes. There are various features of the eye that can be analyzed for ascertaining the eye gaze. These eye-based images are used as input for different application of eye gaze-based systems. But, eye gaze-based systems produced good results if only if these input images are accurate. The outputs of the eye gaze-based systems increase if these digital images are free from any distortion and have minimum noise. Thus, the efficiency of eye gaze-based systems depends upon the image quality. The more the image is clear; the eye gaze-based systems generate good results. Different types of tampering are incorporated with the digital images either manually or by using any image editing softwares. Noise is one of the tampering that may occur due to some resource limitations or intentional addition. This type of tampering disturbs the image basic feature and results in false information to the society. These altered or noisy digital images affect the accuracy of the results. Therefore, it is essential to identify the original image so that the meaningful results can be generated by the eye gaze-based systems using these noise free images [2]. Recently researchers have focused on SVD-based models to remove such type of tampering from digital images. Singular Value Decomposition is one of the robust and efficient methods to produce noise free digital images [3]. This technique involves refactoring of given digital image in three different feature-based matrices. The small set called singular values preserves the useful features of the original image. It has many applications in data analysis, signal processing, pattern recognition, image compression, noise reduction, image blurring, face recognition, forensics, and embedding watermarking to an image [4, 5]. In this research study, SVD-based noise detection and removal (SVD-TR) model [6] has been applied to remove two different noises (salt-pepper and Gaussian) from eye gaze-based image database.

2 Related Work

Different researchers have proposed various methods and techniques of detecting tampering in digital images. Some of the significant research work in this area has been presented below.

SVD-based decomposition tampering detection model transforms the image into different mutually compatible matrices that can express the various relationships among the original data items. The mathematical model allows refactoring a digital image in different segments called singular values, representing a subset, which preserves the useful features of the original image [7]. A versatile denoising method for contaminated digital images with Gaussian Noise has been proposed by Jain et al. Simulation results indicates improvement in the quality of the restored image [8]. A novel approach of decision median filter for suppression of salt and pepper noise in digital images has been presented by Bhateja et al. This algorithm performs decision by comparing the computed median with the minimum and maximum pixel values of image. The technique efficiently suppresses noise contamination levels as

high as 90 % [6]. Sinha et al. describes the concept of image denoising using bilateral filter with rayleigh distribution to reduce the Gaussian noise levels. Simulation-based results on the basis of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) show good results [9]. A novel technique for impulse noise reduction with five different smoothing filters has also been proposed [10–12].

In eye gaze techniques, real-time data is gathered for tracking and estimation of eye gaze in relation to gaze direction of the eye position and movements. The different actions can be recorded based on different eye inputs like blinking, frowning of eye, eyeball movements, view or visual angle, etc. [13]. Different calibrated or non-calibrated eye tracking hardware devices may also induce some noise. The image thus obtained is normalized by performing different kinds of preprocessing and is further analyzed for identification for various parameters [14].

It is evident from the review of different researchers that lot of research has been carried out in the field of eye gaze-based systems and tampering detection. The efficiency of the eye gaze-based systems depends upon the accuracy of the digital images. SVD is one of the robust methods of detecting and removing tampering from digital images. These noise free digital images serve as input to the eye gaze-based systems. The workflow of the SVD-TR model has been discussed in the next section.

3 Proposed Model

After the extensive research survey and literature review of existing techniques of removing tampering from eye gaze-based digital images, it has been observed that there are different types of tampering associated with digital images. The important objective of this research study is the removal of noise from eye gaze-based image database using SVD-TR model. Two types of noise, Salt-pepper and the Gaussian noise, have been induced manually in the eye gaze-based image database. SVD-TR model has been applied and generate a noise free image, which after preprocessing has been used as input to eye gaze-based systems (Fig. 1).

More than 200 eye-based digital images (116 of males and 84 of females) have been collected from captured digital image database using digital camera of SONY NEX-5 ultra compact with resolution 4592×3056 with 14.2 megapixels sensor and a large articulated 7.5 cm monitor specification. This eye gaze-based image database has been evaluated for the presence of noise tampering using the SVD-TR. This model works on the singular values left (SL) and right singular (SR) values of the images. The extent of noise removal from eye gaze-based images depends upon the behavior of the corresponding singular values of the test image and the noisy image. This evaluation of noise content in digital images results in the appropriate function of eye gaze-based systems. Noise has been removed from images so that after preprocessing these images has been used for different applications of eye gaze-based systems for eye gaze estimation and glint detection. The preprocessing process includes the detection and removal of noise from this eye gaze-based image

database. The elaboration of the SVD-TR model has been shown in Fig. 3. SVD-TR model is used to remove the noise from such images and generate noise or distortion free image. The digital eye gaze-based image database is considered to investigate the efficiency and effectiveness of the SVD-TR model. These test images are first normalized in the required format by undergoing cropping, resizing process. In order to evaluate the extent of tampering removal, noise has been added to the eye gaze-based digital images manually. Two standard noises, salt-pepper (mean < 0.5) and Gaussian noise has been incorporated to eye-based digital image database and then the SVD-based noise detection and removal (SVD-TR) model is applied to remove the noise from eye gaze bases image database so that this image database can be further used eye gaze estimation process. The left and the right singular values of the corresponding images have been calculated and then the induced noise has been removed from these images. In order to investigate the extent of noise removal this resultant noise free image is then compared with the original input image. Few of the images from eye gaze-based digital image database are shown in Fig. 2.

4 Methodology

In the present research work, the SVD-based noise detection and removal (SVD-TR) model has been applied as shown in Fig. 3. This system takes the original eye gaze-based image from the database as input and after processing generates the corresponding noise a free image, which has been further, used for the eye gaze-based estimation as shown in Fig. 1. Image I has been taken from the eye gaze-based image database as input to the system. I has been normalized in a required format (I_N).

For testing the SVD-TR model, two standard types of noise, salt-pepper and the Gaussian noise, have been added manually to I . SVD has been applied to I_N in order to extract the basic features of digital image in the form of left (S_L) and right singular (S_R) values, which has been used for further testing. SVD-based detection and removal (SVD-TR) model has been applied to I_N to remove the maximum quantity of noise and generated a resultant noise free image (I_{NF}). I_{NF} has been

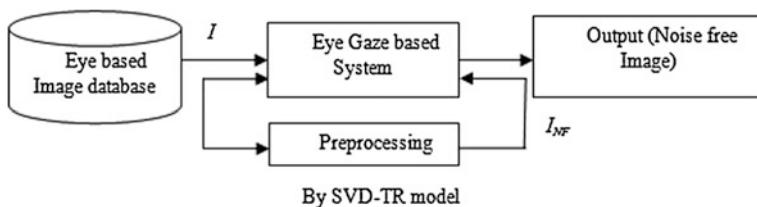


Fig. 1 Normalization process of eye gaze-based image database using SVD-TR model



Fig. 2 Eye gaze-based digital image database

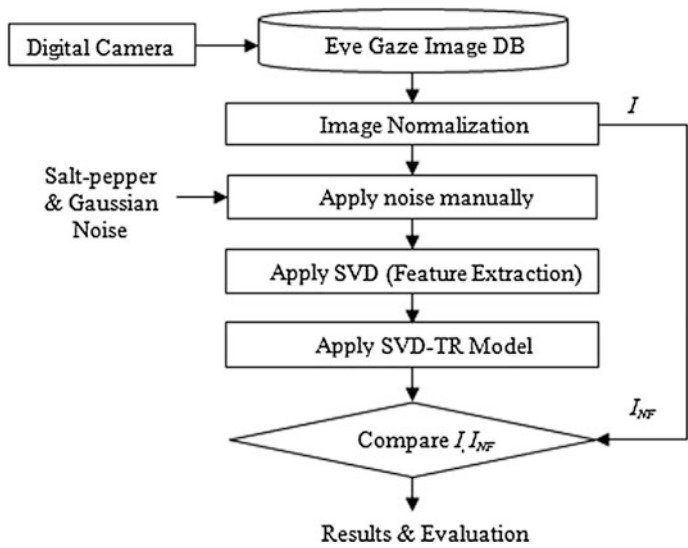


Fig. 3 Schematic diagram of SVD-based noise detection and removal (SVD-TR) model

compared with the corresponding original input image, I to calculate the extent of noise removal and image match.

The experimental analysis of the SVD-TR model has been shown in Fig. 4. After noise removal using SVD-TR model, this eye gaze-based digital resultant image can then be used for eye gaze estimation process as this image is free from any distortion and accurate enough to generate correct and meaningful results. The analysis and the interpretation of the results have been discussed in next section.

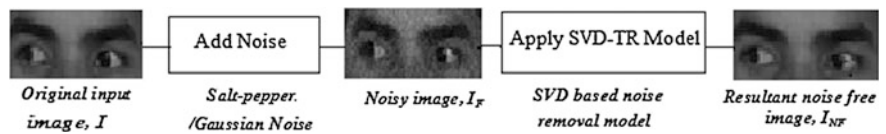


Fig. 4 Experimental analysis of the SVD-TR model

5 Results and Discussions

This section presents the experimental results illustrating the performance of the SVD-based noise detection and removal (SVD-TR) model. The visual outputs of test image gave satisfactory results. In order to check the effectiveness of the model, extent of noise removal has been evaluated by comparing the resultant and the original input image. The variation in the left and right singular values of the original and the corresponding noisy image gives the extent of image tampering. The study of singular values with respect to image tampering has been shown in Table 1.

The singular values (S_L and S_R) of the original images and the resultant image are obtained using the SVD-TR model by computing the SVD of the images in the form of three matrices U , S , and V . As depicted from the table, the singular values of the original image and the corresponding tampered free images are almost similar to each other which indicates that the noise tampering in form of salt-pepper and Gaussian noise has been removed up to great extent. The comparison of the original image and the corresponding noise free image has also been computed by using the pixel difference mean square error (PDMSE). For empirical analysis, the results obtained from the SVD-TR model are then compared with the existing median filtering technique. This comparison in terms of percentage has been shown in Table 2. It is evident from the results obtained by the SVD-TR model and the median filtering that the SVD-TR model works efficiently as compared with the existing filtering technique.

Table 1 Singular value of tampered image and the corresponding noisy image

Images	Noise	Singular values of original images		Singular values after tampering removing	
		S_L	S_R	S_L	S_R
I_1	SP	101.35	0.43	98.73	0.23
	G	128.43	5.92	129.78	3.98
I_2	SP	108.34	0.95	102.93	1.23
	G	103.47	9.38	99.738	7.94
I_3	SP	117.38	0.93	123.91	0.73
	G	128.90	4.91	119.82	6.02
I_4	SP	143.72	7.92	139.09	7.02
	G	136.82	0.74	133.70	1.90
I_5	SP	127.87	1.84	121.93	2.91
	G	109.75	3.98	102.72	2.90
I_6	SP	113.72	0.44	115.92	1.58
	G	114.83	1.93	100.73	1.50
I_7	SP	129.45	4.87	134.91	3.90
	G	110.85	1.03	104.82	2.94

Table 2 Extent of noise removal using SVD-TR model and median filtering (in percentage)

Images	Noise added	Median filter	PDMSE	SVD-TR
I_1	SP	67.75	73.24	87.67
	G	64.89	71.43	78.72
I_2	SP	70.34	89.56	93.41
	G	68.73	65.34	80.24
I_3	SP	69.93	88.34	91.59
	G	66.11	77.32	79.25
I_4	SP	71.82	83.92	87.67
	G	70.72	75.33	77.21
I_5	SP	63.17	84.89	90.25
	G	62.66	70.38	79.24
I_6	SP	63.10	84.88	88.21
	G	60.18	79.11	80.12
I_7	SP	59.19	84.23	89.21
	G	62.01	70.27	77.24

It is observed that salt-pepper noise from the eye gaze-based digital images has been removed up to a great extent, i.e., 90 % whereas Gaussian noise removed from test images lies between 72 and 85 %. This indicates that the SVD-TR model works efficiently for salt-pepper noise as compare with the median filtering technique. Gaussian noise also gave satisfactory results but with little variation.

The investigation indicates that the digital images obtained after removing noise by using the SVD-TR model can be used for eye gaze-based systems for gaze estimation. The results obtained after removing salt-pepper and Gaussian noise using SVD-TR model has been shown in Figs. 5 and 6, respectively. The result shows the extent of tampering removal from the eye gaze-based image database (in % age).

For evaluating the efficiency of the SVD-TR model, the results obtained has been compared with the median filtering and PDMSE technique. The results indicate that the SVD-TR model works efficiently and removes the noise from eye

Fig. 5 Extent of salt-pepper noise removal`

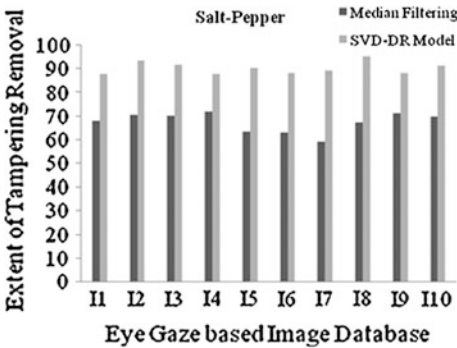


Fig. 6 Extent of Gaussian noise removal

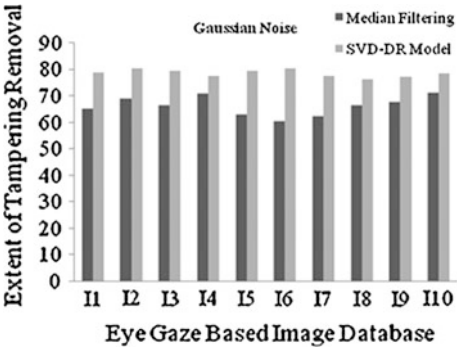
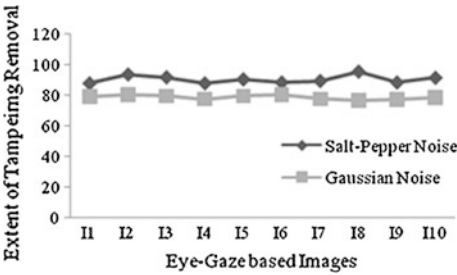


Fig. 7 Comparative analysis of salt-pepper and Gaussian noise using SVD-TR model



gaze-based digital images satisfactory as compared with the results obtained from median filter. The analysis of the results obtained from SVD-TR indicates that the salt-pepper noise has been approximately removed from eye gaze-based digital images as compared with the Gaussian noise. The comparative analysis of both the noise removed from the eye gaze-based images using the SVD-TR model has been shown in Fig. 7.

6 Conclusion

In this research study, SVD-TR model has been applied for detecting and removing the salt-pepper and gaussian noise from eye gaze-based image database. The analysis of the results obtained from the research study indicates that the noise has been removed up to great extent from the eye gaze-based digital image database using SVD-TR model. Tampering in the form of standard salt-pepper noise shows a great extent of removal as compared with Gaussian noise. The SVD-TR model works on the computation of singular values. As compared with the median filtering technique, this model results in the approximate removal of noise from eye gaze-based digital image database. In future, this noise free eye gaze image database can be used as input to different applications of eye gaze estimation process.

References

1. Kim, K., Ramakrishna, R.S.: Vision-Based Eye-Gaze Tracking for Human Computer Interface. In: Proceedings of IEEE Systems, Man and Cybernetics Conference, pp. 324–329 (1999).
2. Li, X. H., Zhao, Y. Q., Liao, M., Shih, F. Y., Shi, Y. Q.: Detection of tampered region for JPEG images by using mode-based first digit features. *EURASIP J. on Advances in Signal Processing*, pp. 1–10 (2012).
3. Wack, D. S., Badgaiyan, R. D.: Complex Singular Value Decomposition Based Noise Reduction of Dynamic PET Images. *Current Medical Imaging Reviews* ©2011 Bentham Science Publishers Ltd., pp. 113–117 (2011).
4. Osmanli, O. N.: A Singular Value Decomposition Approach for Recommendations Systems. M.Sc. thesis, Dept. of Computer Engineering, Middle East Technical University (2010).
5. Wall, M. E., Rechtsteiner, A., Rocha, L. M.: Singular Value Decomposition and Principal Component Analysis. *J.A Practical Approach to Microarray Data Analysis*, Computer and computational division, pp. 91–109 (2003).
6. Sinha, A. K., Bhateja, V., Sharma, A.: Rayleigh based Bilateral Filtering for Flash and No-Flash Image Pairs. (IEEE) 2nd International Conference on Signal Processing and Integrated Networks (SPIN-2015), pp. 559–563, February (2015).
7. Gul, G., Avcibas, I., Kurugollu, F.: SVD based image manipulations detection. In: 17th Int. Conf. Image Processing, pp. 1765–1768 (2010).
8. Jain, A., Bhateja, V.: A Versatile Denoising Method for Images Contaminated with Gaussian Noise. In: (ACM ICPS) CUBE International Information Technology Conference & Exhibition, pp. 65–68 (2012).
9. Bhateja, V., Verma, A., Rastogi, K., Malhotra, C., Satapathy, S. C.: Performance Improvement of Decision Median Filter for Suppression of Salt and Pepper Noise. *Proc. (Springer) International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS-2014)*, vol. 264, pp. 287–297, March (2014).
10. Harikiran, J., Divakar, B.: Impulse Noise Removal in Digital Images. *Int. J. of computer applications*, vol.10 no. 8, pp. 39–42 (2010).
11. Sharma, D., Abrol, P.: SVD Based Noise Removal Technique: An Experimental Analysis. *International J. of Advanced Research in Computer Science*, pp. 214–218, vol. 3, no. 5, Sept-Oct (2012).
12. Lysaker, M., Lundervold, A., Tai, X.: Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE transactions on Image Processing*, vol. 12, no. 12, pp. 1579–1590, December (2003).
13. Poole, A., Ball, L. J.: Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. *Encyclopedia of Human-Computer Interaction*, Idea Group Inc. (2005).
14. Sharma, A., Abrol, P.: Eye Gaze Techniques for Human Computer Interaction: A Research Survey. *International J. of Computer Applications* vol. 71, no. 9, pp. 18–29 (2013).

Identity-Based Key Management

Purvi Ramanuj and J.S. Shah

Abstract Mobile ad hoc networks (MANETs) are more vulnerable to security attacks compared to the wired networks mainly because they are wireless and dynamic in topology. It becomes very crucial to provide secured and efficient key management scheme as well as all the messages should also be secured. We hereby propose a scheme which provides secure identity-based key management which includes key generation using finger print data as identity of user and Key revocation. The proposed scheme reduces load on network and required computational time at receiver end by sending modified revocation list in accusation messages. Instead of sending entire accusation list, only changes in accusation list are sent. Also, any previous accusation from the current revoked node will be discarded. This results into increased efficiency and enhanced performance of system.

Keywords Mobile ad hoc network • ID-based security • Key management

1 Introduction

Mobile ad hoc network (MANET) is an infrastructure-less collection of mobile devices connected by wireless links. MANET nodes are free to move from one place to another and also change their configuration dynamically. Each node functions as both a host and a router. Nodes leave the network, join the network or change their positions dynamically so there is no fixed network topology it is dynamic. There is no fixed communication structure and no base station is present to organize communication pattern.

Purvi Ramanuj (✉)
R K School of Technology, R. K. University, Rajkot, India
e-mail: ramanuj.pn@gmail.com

J.S. Shah
Gujarat Institute of Technical Studies, Himmatnagar, India
e-mail: jssld@yahoo.com

1.1 Security in MANET

Security represents one of the most important issues in communication between networks. Also due to its inherent nature, as discussed above, it is more crucial to secure networks in MANET environment. Routing plays an important role in providing security. A secured routing protocol needs to protect the sessions from intruders or from any other illegal operations. Key management plays an important role to establish secured communication. A comprehensive, efficient, robust, and secured key management scheme is very crucial for secure communication. Various key management schemes in MANET are available in [1] and can be classified as in (Fig. 1).

1.2 Identity-Based Cryptosystem

Shamir [2] introduced a unique combination of identity and cryptosystem for the first time. Various identities like email, IP address, or any biometric property is used to generate public key and a trusted third party (PKG) generates the private key. Such a combination of identity and cryptography was known as ID-based cryptography (IBC). But the first efficient scheme was introduced by Boneh and Franklin [3]. IBC have upper hand compared to traditional public key methods in term of simplicity in key management, reduced system resources like memory storage and computational power. Only PKG parameters, and not the entire public key certificates, are needed in IBC which makes it very easy to deploy, very less

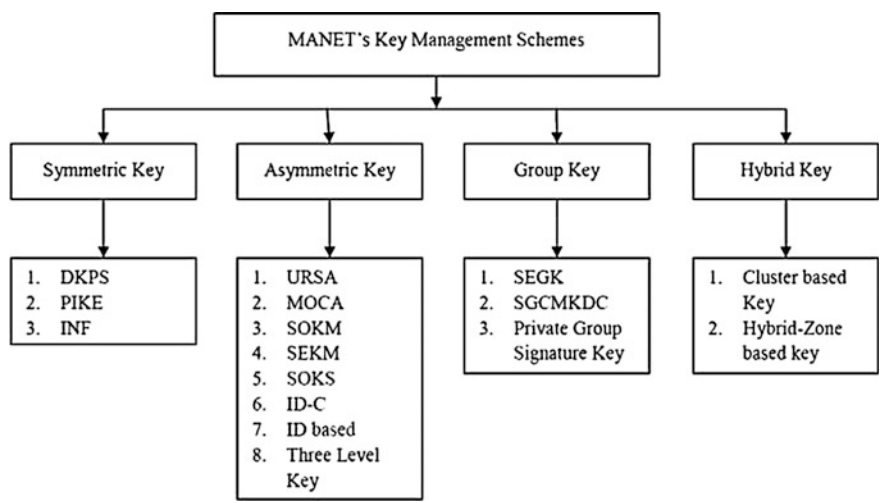


Fig. 1 Key management schemes in MANET

requirements on infrastructure, and huge saving on certificate distribution. IBC is also efficient on various system resources requirements like processing power, storage space, and communication bandwidth.

The public key of IBC is self-proving and can carry much useful information, provides authentication, confidentiality, non repudiation, and integrity. Here, we have proposed an ID-based key management scheme which includes key generation and key revocation.

2 Preliminary

In this section, we briefly describe the basic technology on which our scheme is based on.

Choosing identities based on the needs and application requirements, various strings or identity can be chosen. We need to focus on whom we want to be identified or authenticated in the network. Generally, we can distinguish three cases of entities like user operating a network node, a node itself or network interface of a node.

Various identities like e-mail ID, name, IP Address, etc., are used for identity of a user. Also, biometrics like finger prints, retina, iris, face, signature, voice, hand geometry, etc., are suitable candidates for identity of a node. Any identity used in ID-based framework must satisfy the following properties: Unique for each entity in the network, unchangeably bound to an entity for its entire lifetime and non transferable. We have preferred biometrics for enhanced security. Out of various biometrics, finger print is one of the oldest candidates. Finger print details of each user is captured, digitized and stored at KGC. At the time of network initialization or when a node joins a network, it requires providing its finger print details. KGC will compare the received finger print details with that of the stored values and thus authenticate the user.

Because of weak physical protection of nodes and node's exposure to potentially hostile environments, node compromises as well as key disclosures are very obvious in MANETs. Hence, key revocation and key renewal are of great importance in MANETs. There are various revocation schemes that have been especially designed for MANETs, e.g., [4–12], but they either completely ignore key revocation and/or key renewal or just says a little about possible solutions. Many of these schemes require presence of KGC throughout the network operation which is not feasible for MANETs. They required for computation of trust which is computationally very demanding. Also the propagation of observations is also not satisfactory and requires more computational power and demands large number of revoked keys. As per [12], a separate KRL–Key Revocation List is maintained at each node which registers the trust values of each m-hop node. A node will monitor behavior of its immediate neighboring node and records observations in KRL. For other than immediate neighbor, the observation from other nodes will be received and as per majority of received observations, the status in KRL is updated.

But in this scheme, entire KRL is being sent to all neighboring nodes. Also, accusation message update is sent every time status change is observed even for one of the node. This requires more computational power and storage for KRL update and processing accusation message.

For detection of malicious node, we require a metric to measure malicious behavior, a scheme to observe the specified behavior and a scheme to punish identified nodes. Due to the lack of a central TTP, identifying and excluding/punishing malicious nodes must be carried out by network nodes themselves. Node's malicious behavior can be identified by any or combination of: number of dropped packets, number of generated packets, measuring response time of nodes wait for messages confirming each hop on a multi hop routing path, use of suitable anomaly detection systems for detection of unusual behavior or running Intrusion Detection System on each node for identifying so-called signatures of known attacks. Depending on the measures selected for detection, it requires monitoring of neighboring nodes, observation of routing behavior, complex behavior patterns, or even running special software on nodes. Out of these schemes, we choose monitoring of neighboring nodes behavior as it is more economic and efficient. Also, it has been assumed that such kind of observation mechanism is already built up in the nodes.

3 Proposed Scheme

Proposed scheme for the comprehensive solution can be distributed into the following two major parts: Key Generation and Key Revocation.

Key Generation: Private Key Generator (PKG) will generate its own master public key and master private key using RSA algorithm (As per Shamir's scheme). Nodes will provide finger print data to PKG to establish identity. PKG will generate Public/Private Key of the node using Elliptic Curve. This key will be used for encryption and authentication messages. PKG provides a private and public key pair ($d_i; Q_i$) to each node prior joining the network. The public key format is as given below

$$Q_i(t_x; v_i) = H1(ID_i || t_x || v_i) \quad (1)$$

where ID_i is the identity of node i , t_x is expiration time and V_i is version of the key. This allows key to be renewed after fixed time interval as well as it allows node to ask for renew the key at any time. By this way, we can achieve tradeoff between user friendliness and performance.

Key Revocation: Our revocation scheme is based on the following assumptions: the communication links are bidirectional, a suitable monitoring scheme is already implemented on all nodes, each node i is identifiable by a unique identity ID_i , each node is aware about identities of all the nodes in m -hope as well as distance of all nodes in immediate neighborhood is known.

In the proposed scheme, neighbor node observation algorithm is used by each node to find out any suspicious behavior. It is used for monitoring the nodes only in a node's direct communication range. Any malicious observation is recorded then the node will mark it as a malicious node. Also such observation will be securely propagated to the m -hop neighborhood. For any node which is in m -hop proximity but not in direct communication range, it requires at least σ nodes to accuse it for revocation of its corresponding public key. Parameters m and σ can be adjusted as per the level of security needed and performance asked for. Each node maintains node key list—NKL where ID of a directly accused node or a reported accused node will be stored. Thus, NKL will be vector maintaining IDs of revoked nodes. Any additions in NKL will be propagated to neighboring nodes.

The proposed key revocation scheme has the following algorithms:

Neighbor node observation—a node will monitor the nodes in its one hop neighborhood for suspicious behavior. For every expiry period, the node sets accusation values for its neighbors and if any node has been observed as malicious from trustworthy, a neighbor node accusation message as below is propagated:

$$NAM_{i,j} = (fKi; j(ID_i; nmi); (ID_i; nmi)) \quad (2)$$

for all nodes which were marked as suspicious. This messages is secured by a MAC function f .

Propagate—Accusations are securely sent to all neighbors. *Update NKL*—nodes update their neighbor key lists using received accusations either neighbor node accusation message or an NKL update message. If the accumulated accusation against a node exceeds a certain value, the key for that node will be revoked.

3.1 Performance Analysis

Performance of the revocation scheme depends how malicious the network is, i.e., the frequency of accusation messages are sent, or how many malicious nodes are there in the network. Also, the frequency and trigger for sending accusation message is also important. We can think of two mechanisms for sending accusation messages. One, to send accusation messages as and when malicious behavior observed during neighbor node observation or after NKL update. Second, accusations are propagated periodically. In the first case, accusations will travel fast but the communication overhead will be high. While if the rate of malicious nodes is higher in the network, then second approach, will be useful. Another parameter that affects the network performance is selection of m , i.e., propagation range. We can adjust m accordingly as discussed earlier. The smaller m , the lesser a message must be resent to the next hop. Smaller m decreases the network load but on the other hand it will restrict number of nodes to which communication is possible with trust details.

In the earlier scheme [12], each node is maintaining NKL for all the nodes in the network. This will include the node’s own observations, i.e., neighborhood watch as well as the other accusations received from the other nodes in the network. So if there are n nodes in the m -hope, the size of NKL matrix will be $n \times n$. We can see that the size increases exponentially with increase in nodes. This will require more storage space and more computational power for the nodes to maintain NKL. The proposed scheme enhances the performance of the scheme by maintaining only a revocation list by each node. Revocation list contains only accusations against each m -hope neighboring nodes. If a node is not yet accused by any node in the network, it will not appear into this list. So in this scheme, the NKL is of size $n \times 2$ instead of $n \times n$ in the earlier scheme. The two columns will be node id and the number of accusations.

4 Implementation and Results

We have simulated the proposed scheme in Java and MATLAB. Key generation has been implemented in MATLAB and master public key and private key are generated using RSA. Nodes public key and private key are generated using master public key and portion of finger print data. For the purpose of key revocation we have created network scenario with different no of nodes in network. We have also varied the number of source nodes. Time taken by nodes to update its NKL using existing scheme [12] and the proposed algorithms is measured. Scenario 1: Execution time comparison to update key revocation list for network consist of 45 nodes with different no of source nodes (Fig. 2, Table 1).

Scenario 2: Execution time comparison to update key revocation list for network consist of 55 nodes with different no of source nodes (Fig. 3, Table 2).

Fig. 2 Comparison of KRL/NKL update time in milliseconds

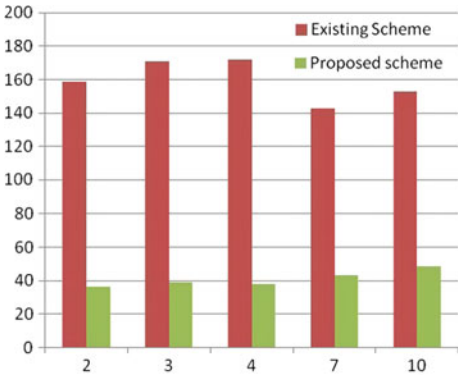


Table 1 Comparison of execution time to update KRL/NKL in milliseconds

No of source nodes	KRL update	NKL update
2	148.44	34.31
3	165.73	37.21
5	167.67	37.29
8	141.88	42.23
10	155.77	44.43

Total nodes in network = 45

Fig. 3 Comparison of KRL/NKL updates time in milliseconds

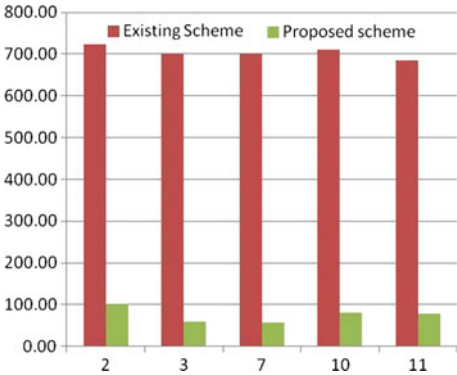


Table 2 Comparison of execution time to update KRL/NKL in milliseconds

No of source nodes	KRL update	NKL update
2	713.61	100.55
3	688.92	56.86
5	689.41	55.94
8	700.73	79.18
10	674.39	76.78

Total nodes in network = 55

Results clearly indicates the improvement in the performance as we have reduced the NKL which contains details of only accused nodes and not all the m-hope neighbors as in the earlier scheme. A significant improvement in NKL update time has been noted.

5 Conclusion

A comprehensive security can be provided using ID-based cryptography and using proposed key management scheme. The proposed scheme provides a significant improvement in the performance of key revocation scheme. Reduced size of NKL helps in faster propagation and processing of observation messages. Apart from the

current reduced size of NKL, we can also apply reactive approach in NKL update, i.e., any node asks for trust observations only when the node has data to send to the other node. The same can be clubbed with the routing information also. This will be our future enhancement in the proposed scheme.

References

1. Renu Dalal, Yudhvir Singh and Manju Khari, A Review on Key Management Schemes in MANET, International Journal of Distributed and Parallel Systems (IJDPs) Vol. 3, No. 4, July 2012.
2. Adi Shamir. Identity-based cryptosystems and signature schemes. In Proceedings of CRYPTO 84 on Advances in cryptology, pages 47–53. Springer-Verlag New York, Inc., 1985.
3. Dan Boneh and Matthew K. Franklin. Identity-based encryption from the weil pairing. In Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology, pages 213–229. Springer-Verlag, 2001.
4. J. Clulow and T. Moore. Suicide for the Common Good: a New Strategy for Credential Revocation in Self-organized Systems, ACM SIDOPS Operating Systems Reviews, vol. 40, no. 3, pp. 18–21, 2006.
5. C. Crepeau and C.R. Davis. A Certificate Revocation Scheme for Wireless Ad Hoc Networks, Proceedings of ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN '03), ACM Press, ISBN: 1-58113-783-4, pp. 54–61, 2003.
6. H. Deng, A. Mukherjee, D.P. Agrawal. Threshold and Identity-based Key Management and Authentication for Wireless Ad Hoc Networks, International Conference on Information Technology: Coding and computing (ITCC'04), vol. 1, pp. 107–115, 2004.
7. J.P. Hubaux, L. Buttyan and S. Capkun. The Quest for Security in Mobile Ad Hoc Networks, ACM Symposium on Mobile Networking and Computing {MobiHOC 2001, 2001.
8. A. Khalili, J. Katz, and W. Arbaugh. Toward Secure Key Distribution in Truly Ad-Hoc Networks, 2003 Symposium on Applications and the Internet Workshops (SAINT 2003), IEEE Computer Society, ISBN: 0-7695-1873-7, pp. 342–346, 2003.
9. H. Luo, P. Zerfos, J. Kong, S. Lu, and L. Zhang. Self-Securing Ad Hoc Wireless Networks, Seventh IEEE Symposium on Computers and Communications (ISCC '02), 2002.
10. Y. Zhang, W. Liu, W. Lou, and Y. Fang. Securing Mobile Ad Hoc Networks with Certificate less Public Keys. IEEE Trans. Dependable Secure. Computing, vol. 3, no. 4, pp. 386–399, 2006.
11. L. Zhou and Z.J. Haas. Securing Ad Hoc Networks, IEEE Network Journal, vol. 13, no. 6, pp. 24–30, 1999.
12. Katrin Hoeper and Guang Gong “Key Revocation for Identity-Based Schemes in Mobile Ad Hoc Networks”, Ad-Hoc, Mobile, and Wireless Networks, 2007 – Springer.

Firefly Algorithm Hybridized with Flower Pollination Algorithm for Multimodal Functions

Shifali Kalra and Sankalap Arora

Abstract The successful evolutionary characteristics of biological systems have motivated the researchers to use various nature-inspired algorithms to solve various real-world problems that are complex in nature. These algorithms have the capability to find optimum solutions faster than conventional algorithms. The proposed algorithm uses two terms, exploration and exploitation, effectively from Firefly Algorithm (FA) and Flower Pollination Algorithm (FPA). The proposed algorithm (FA/FPA) is validated using various standard benchmark functions and further its comparison is done with FA and FPA. The result evaluation of the proposed algorithm compute better performance than FA and FPA on most of the benchmark functions.

Keywords Swarm intelligence · Firefly Algorithm · Flower Pollination Algorithm · Hybridization

1 Introduction

There is a rapid evolution in the algorithms from the past few decades that are inspired from natural behavior of biological species, which are based on certain successful characteristics of biological system. These algorithms mimic the social behavior of species like birds, bees, ants, etc; so-called Swarm Intelligence (SI) [1]. The reason for their popularity lies in their ability to solve real-world global optimization problems efficiently. Among these biology-derived algorithms, various swarm intelligence algorithms have been proposed till date like ant colony optimization (ACO) [2], particle swarm optimization (PSO) [3], firefly algorithm

Shifali Kalra (✉) · Sankalap Arora
DAV University, Jalandhar, India
e-mail: Kalra_shifali@yahoo.com

Sankalap Arora
e-mail: Sankalap.arora@gmail.com

(FA) [4], multiobjective flower pollination algorithm (MPFA) [5], etc. These various SI algorithms are inspired from simple concepts which relates to physical phenomenon and animal behaviors [6]. These algorithms have received remarkable attention as they are known to be derivative free, robust, and can be applicable to different optimization problems [7]. These algorithms use the concept of randomization that has its efficiency moving away from local search to the global search [8].

However, applying such single methods to solve optimization problem is not very effective, due to their slow convergence rate. This is because these methods usually require a huge amount of computational times and they get frequently trapped within local search space. So, to increase the benefits of the optimization algorithms, various optimization algorithms are being combined to compute better outcomes. These optimization algorithms which contain good features of two or three algorithms have proved their effectiveness in terms of computational time and convergence rate [9].

In this research paper, the proposed algorithm is based on two metaheuristic algorithms that are: Firefly Algorithm and Flower Pollination Algorithm that are thoroughly investigated [10]. This proposed algorithm is a combination of both these algorithms that uses the concept of exploration (diversification) and exploitation (intensification) terms in its algorithm. The exploration term act as a global search and exploitation term act as a local search [7]. The proposed algorithm will be compared with FA and FPA based on their performance. These three algorithms will be tested on various standard benchmark functions and their performance will be evaluated based on parameters that are: Convergence rate and time consumption.

This paper is organized in following manner; Section 2 provides the review on the different optimization algorithms. Then, FA and FPA algorithms are elaborated, respectively, in Sects. 3 and 4, respectively. Then, Sect. 5 presents the proposed algorithm. Section 6 discusses the simulations that are performed on various benchmark functions. Then, Sect. 7 discusses the comparisons and results between PA, FA and FPA. At last, Sect. 8 will include conclusion and some future scope.

2 Related Work

This section gives the brief review about the different new biologically inspired swarm intelligent algorithms. The various swarm intelligent algorithms described below and they have been widely used to solve various optimization problems.

Ant colony algorithm (ACO) is inspired from the foraging behavior of ants, i.e., it is based on the self-organizing behavior of ants [2]. The ants search out a shortest path from their colony to food sources. The ants communicate using a volatile chemical substance called pheromone. The path selection is made by other ants by lying down the pheromone trails, providing positive feedback. This phenomenon of high coordination among real ants to find shortest path can be exploited further to

make a coordination between various artificial agents that collaborates to solve an optimization problem [2].

Particle Swarm Optimization (PSO) was introduced by Kennedy and Eberhart in 1995, which is motivated from swarm behavior of fishes, birds schooling [3]. In PSO, each solution acts as a ‘bird’ in the flock that referred to as a “particle.” Particles mimic the natural phenomenon of flock of birds that coordinate together when they fly. The bird with its best location is identified by the flock. Every bird moves toward the best bird having its velocity that is based on its current position. Every bird then searches the space from their new local position, and the process repeats until the flock reaches its desired position [11].

ABC is inspired from the intelligent foraging behavior of bees namely: employed bees, onlooker and scout [12]. The abandoned food source is determined and is exchanged with the new food source located by scouts [13].

Firefly Algorithm (FA) is inspired from the natural behavior of fireflies and their bioluminescence phenomenon, i.e., based upon flashing pattern of fireflies [4]. These fireflies move toward attractive firefly that will act as the current global best one. The flashing light of fireflies is calculated with the help of optimization [7]. FA is used to solve various optimization problems. For example, Traveling Salesman Problem (TSP) using discrete distance among two fireflies and the movement of fireflies [12, 14].

Flower Pollination Algorithm (FPA) is based on the characteristics of flowers of different plants. The main motive is ultimately reproduction through transferring of pollens, and pollinators help in their transfer like insects, birds, bees, and flies [5]. There are two types of pollination which are (1) Abiotic (self-pollination) (2) Biotic (cross-pollination) pollination. Global pollination (Biotic) occurs for long distance, because pollinators for a fly a long distance. Cross-pollination occurs within the flowers of same plants. Both these processes are controlled by a switch probability p to achieve optima faster [10].

3 Firefly Algorithm

3.1 Firefly Algorithm and Its Detailed Concept

The functioning of fireflies is a procedure which is based on flashing light of fireflies, i.e., produced by a natural process called bioluminescence [4]. Firefly emits light in order to attract potential prey and also to attract mating partners. To make them unappetizing to predators, fireflies produce defensive steroids from their bodies. Therefore, the flashing lights are protecting the fireflies from their enemies [4, 15].

The firefly acts as a premium light source and emits some light intensity at some distance r and follows inverse square law. The light intensity I decreases with increase in distance r defined as $I \propto \frac{1}{r^2}$. Air acts as an absorbent medium that has the

capability and absorbs the light in the medium that decreases the visibility of fireflies at some distance [8]. FA is based upon three rules that are:

1. All fireflies are of single sex so the less bright firefly will be attracted toward brighter firefly despite of their sex.
2. The attractiveness between two fireflies is comparable to their brightness.
3. Brightness of the firefly is determined by the objective function.

3.2 *Light Intensity and Attractiveness*

Firefly Algorithm is based on two important things: (1) Modifications in light intensity and (2) Construction of attractiveness. For simplicity, the attractiveness of firefly is examined by its brightness which is further related to the objective function. The brightness of firefly at particular position is written as $I(x) = f(x)$. The attractiveness β will change with distance r_{ij} between firefly i and firefly j [16].

As the distance increases, the brightness of the firefly decreases because light is absorbed in media like air, rain, etc. Therefore, light intensity decreases with increase in the distance from the source. So, the light intensity $I(r)$ varies according to inverse square law as given as $I(r) = \frac{I_s}{r^2}$.

I_s is the source intensity. The light intensity changes with the distance r for a particular light absorption coefficient γ , i.e.,

$$I = I_0 e^{-\gamma r} \quad (1)$$

As mentioned in Eq. (1), I_0 is for initial light intensity. In order to neglect singularity at $r = 0$ in the expression $\frac{I_s}{r^2}$, the effects of both the inverse square law and absorption have been combined and can be approximated using Gaussian form in Eq. (2)

$$I = I_0 e^{-\gamma r^2} \quad (2)$$

In FA, attractiveness β directly corresponds to the light intensity and visualized by other fireflies and is defined in Eq. (3)

$$\beta = \beta_0 e^{-\gamma r^2} \quad (3)$$

The distance between two fireflies is computed using Cartesian distance method as shown in Eq. (4)

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (4)$$

Firefly i is attracted to brighter firefly j and its movement is determined by Eq. (5)

$$x_i = x_i + \beta o e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha (\text{rand} - \frac{1}{2}) \quad (5)$$

As in Eq. (5), second component is due to attraction and third term α is randomization parameter, and rand is for random numbers whose value is taken with in uniform distribution [16]. The parameter γ is air absorption coefficient used to determine the speed of convergence. Its value lies between $\gamma \in [0, \infty)$ and for most of the applications it varies from 0.1 to 10 [4, 17].

3.3 Pseudo Code of Firefly Algorithm

```

1. Define an Objective function min or max  $f(x)$ ,  $x = (x_1, x_2, \dots, x_d)$ 
2. The initial population of fireflies is computed  $x_i$  ( $i = 1, 2, \dots, n$ )
3. The light intensity  $I$  is calculated that is in direct proportionate to objective function  $f(x)$ 
4. Define light absorption coefficient  $\gamma$ 
5. while ( $t < \text{Max-iterations}$ )
    for  $i = 1: n$  (all  $n$  fireflies)
        for  $j = 1: i$  (all  $n$  fireflies)
            if ( $I_i > I_j$ )
                Move less brighter firefly  $i$  towards brighter firefly  $j$  in all  $d$  dimensions;
                Attractiveness changes with distance  $r$  by means of  $\exp[-\gamma r]$ 
                Calculate new solutions and re-evaluate light intensity
            end if
        end for  $j$ 
    end for  $i$ 
    The fireflies are ranked according to light intensity and discover the current best.
end while

```

4 Flower Pollination Algorithm

4.1 Pollination of Flowering Plants

There are more than a quarter of million categories of flowers and most of the 80 % plant species are flower species. The main motive of flower is ultimately to reproduce through transferring of pollens, and pollinators help in their transfer [18]. So, there are two forms of pollination process which are given as below.

1. Biotic pollination (Cross-pollination or allogamy)

There are 90 % of flowering species that belong to biotic pollination also considered as global pollination which occurs at long distance by pollinators, because pollinators like bees, bats and flies fly a long distance [19]. These behave as Levy flight, and jump steps of these pollinators obey a Levy distribution [19, 20]. Flower constancy is developed by these pollinators that move at a certain plant rather than moving at other flowering plants. This will increase chances of the transferring of

pollen to the same plants [21]. The main motive of the pollination is to survive the best one to the better reproduction of plants in terms of numbers and the best one [19].

2. Abiotic pollination (Self-pollination)

There are 10 % of flowering plants which belongs to abiotic or self-pollination form and these do not require such pollinators [18]. Here, fertilization of one flower takes place within the different flowers of the same plant. Wind, diffusion and grass are taken as pollinators for such kind of pollination of plants [19, 20].

4.2 Flower Pollination Algorithm

Flower pollination algorithm is based on four rules that are:

1. Biotic and cross-pollination is regarded as a global pollination process, and the movement of pollinators that carry pollens obeys Levy flights.
2. Abiotic and self-pollination is being taken for local pollination.
3. Pollinators create flower constancy that is comparable to reproduction probability and is directly proportional to the equivalence between two flowers taken for reproduction.
4. Local pollination and global pollination interaction is adjusted by using a switch probability $p \in [0, 1]$ which is slightly more biased toward local pollination.

So, a flower-based algorithm is created, known as, flower pollination algorithm (FPA) [20]. In global pollination, the reproduction of the fittest one is represented as the most fittest as g^* . The first rule and the flower constancy is mathematically formulated in Eq. (6)

$$x_i^{t+1} = x_i^t + L(x_i^t - g^*) \quad (6)$$

x_i^t is taken for pollen i or solution vector x_i at iteration t , and g^* is for current best solution that is evaluated from all the solutions. The L parameter is taken for the efficiency of the pollination that is usually a step size. As insect fly over a long distance with different distance steps, so the Levy flight is used to mimic this phenomenon effectively. That is, $L > 0$ is taken from a Levy distribution [18].

$$L \sim \frac{\lambda \Gamma(\lambda) \sin(\pi\lambda/2)}{\pi} \frac{1}{s^{1+\lambda}}, (s \gg s_0 \gg 0) \quad (7)$$

$\Gamma(\lambda)$ is taken as the standard gamma function and its distribution is important for large steps $s > 0$. In all the simulations discussed below in section, $\lambda = 1.5$ is being used [21].

The local pollination (Rule 2) and flower constancy are given in Eq. (8).

$$x_i^{t+1} = x_i^t + \epsilon (x_j^t - x_k^t) \quad (8)$$

where x_j^t and x_k^t are pollens taken from the two different flowers of the same plant species. These usually mimic the flower constancy in near about species [19, 20]. Mathematically, x_j^t and x_k^t are taken from the same specie population, as it is considered as a local random walk, if ϵ is taken from a iform distribution in $[0, 1]$.

The switch probability or proximity probability p controls the interaction between biotic pollination and abiotic pollination. Value of p is taken as 0.8 [19–22].

4.3 Pseudo Code of Flower Pollination Algorithm

```

1. Define an objective function min or max  $f(x)$ ,  $x = (x_1, x_2, \dots, x_d)$ 
2. Create the population of  $n$  flowers/pollen gametes having random solutions.
3. To search for best solution  $g_*$  in the initial population
4. Determine a switch probability  $p \in [0, 1]$ 
5. Describe a stopping condition (with a fixed number of iterations or accuracy)
   while ( $t < \text{Max-iterations}$ )
     for  $i = 1 : n$  (all  $n$  flowers in the population)
       if ( $\text{rand} < p$ ),
         Define a ( $d$ -dimensional) step vector  $L$  that obeys the concept of Levi distribution
         Do Global pollination  $x_i = x_i^t + L \cdot (g_* - x_i^t)$ 
       else
         Draw  $\epsilon$  from a uniform distribution in  $[0, 1]$ 
         Do local pollination  $x_i^{t+1} = x_i^t + \epsilon (x_j^t - x_k^t)$ 
       end if
       Generate new solutions
       If new solutions are superior, upgrade them in the population
     end for
     Find the current best one pollen or solution  $g_*$ 
   end while
   Output the best solution evaluated

```

5 The Proposed Algorithm (FA/FPA)

This proposed algorithm uses the biological concepts of both algorithms FA and FPA. The main motive of hybridization is to overcome the disadvantages of existing individual components of optimization algorithms and to achieve an improved form. Second, to determine the strength of this proposed algorithm to attain global optima in a short period of time as much as possible and to use the concept of exploration and exploitation efficiently. Both these terms are used to explore the new possible outcomes and to intensify the current solution to make it more superior.

5.1 Basic Concept of Proposed Algorithm (FA/FPA)

The proposed algorithm introduces the concept of both these algorithms. In this algorithm, the movement of particles is based on movement of less brighter particle toward brighter one by performing the local walk and global walk in two steps some kind of having similarities of Firefly Algorithm and Flower Pollination Algorithm. This FA/FPA algorithm includes the concept of Firefly Algorithm firstly by performing local search because all particles make several subgroups and then they found best one value from each group. From all these values, they found a global best one value by avoiding trapping within local optima and decreasing the randomization effect so that particles will be able to explore a better optima solution. So all this process includes the global step, i.e., performed by particles effectively. Then to make an interaction among local and global search, this proposed algorithm (FA/FPA) uses the concept of Flower Pollination Algorithm. Therefore, a switch probability is defined whose value will be greater than the random number generation of particles because in this case the particles move in any direction so the effect of randomization is more in case of local walk.

The local search of Flower Pollination Algorithm includes the exploitation effect as the flowers of same species are being chosen by performing flower consistency process and transfer of pollens is performed on same plant. Similarly, local search of proposed algorithm perform exploitation among particles of same species of particles. So the convergence rate will be fast because particles will perform a local search more efficiently.

5.2 Pseudo Code of the Proposed Algorithm (FA/FPA)

```

1. Define an objective function  $f(x)$  for minimization and maximization problem
2. Start with initializing a new population of particles in search space i.e.  $x_i (i = 1, 2 \dots n)$ 
3. Light intensity at  $x_i$  is evaluated by objective function  $f(x_i)$ 
4. Determine a air absorption parameter i.e. light absorption coefficient  $\gamma$ 
5. Determine a switch probability whose value lie with in uniform distribution i.e. 0 and 1
6. while ( $t < \text{MaximumGenerations}$ )
    for  $i = 1 : n$  (all  $n$  particles)
        for  $j = 1 : i$  (all  $n$  particles)
            if ( $I_i > I_j$ ),
                The less bright  $i$  particle move towards brighter  $j$  more bright agent in all  $d$ -dimensions.
                if ( $\text{rand} < p$ )
                    The Attraction between fireflies varies with distance  $r$  via  $\exp [-\gamma r^2]$ 
                    Do Global walk via

$$x_i^{k+1} = x_i^k + \left( \frac{\text{beta}}{r} \right) * (x_j^k - x_i^k) * (r * (rt)^2 - \text{tmpf})$$

                else
                    Do Local Walk via

$$x_i^{k+1} = x_i^k + \left( (x_i^k - x_i^k) * (r1)^2 \right)$$

            end if
        end if
    end for
    Evaluate and update new solutions.
end for  $j : \text{for all } n \text{ agents}$ 
end for  $i : \text{for all } n \text{ agents}$ 
7. Set the position of the agents by ranking them and find the current global best particle.

```


Attractiveness parameter is computed in Eq. (9), i.e., the movement of the particle i toward j is given as

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (9)$$

whereas, β_0 is the attractiveness when the value of $r = 0$ and γ is an absorption coefficient parameter. The distance among particles is calculated by the Cartesian Distance is given as in Eq. (10)

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (10)$$

where, $x_{i,k}$ is the k th component of the spatial coordinate x_i of the i th particle.

The global walk is mathematically formulated as given below Eq. (11)

$$x_i^{k+1} = x_i^k + (\text{beta}/r) * (x_j^k - x_i^k) * (r * (rt)^2) - \text{tmpf} \quad (11)$$

where x_i^k and x_j^k are agents or solution vectors at iteration k and movement of i particle is attracted toward brighter one j . So for attractiveness, beta/r involves the decrease in the attractiveness with increase in the distance r between particles as they move far away from each other. So, the light will be absorbed by the medium due to which attraction becomes weak. The term rt is taken for random number generator of particles between 0 and 1. The tmpf parameter is taken to decrease the effect of randomization. The tmpf parameter is computed mathematically as given in Eq. (12)

$$\text{tmpf} = \alpha * \left(r1 - \frac{1}{2} \right) * \text{scale} \quad (12)$$

The value of α is taken for randomization to decrease its effect. The $r1$ parameter is taken for generating random number values of particles between 0 and 1. The scale parameter is taken for the value whose range lie within the domain of any benchmark function.

The local walk is mathematically formulated as Eq. (13)

$$x_i^{k+1} = x_i^k + \left((x_l^k - x_i^k) * (\text{rand})^2 \right) \quad (13)$$

where, l and rand parameters are taken to increase the effect of randomization. Where l contains the value of random index of the particle and rand is taken as a random number generator whose value will lie between 0 and 1.

However, all these three algorithms are efficient in their own way but comparison shows that there are still some factors that need to be improved in case of FA and FPA like convergence rate as they take large amount of time to explore the solution and get easily trapped within local optima. The optimal solution is good

but to obtain further better optimal solution and this new proposed algorithm is proposed that converges faster and decreases the computational time.

6 Benchmark Functions and Experimental setup

This new proposed algorithm is validated by making a comparison with other standardized algorithms. The comparison is made on the basis of various ways. The approach used in this algorithm is on the basis of accuracy of the evaluation of the benchmark functions. So, for its validation, the new proposed algorithm is validated on ten benchmark functions and is given above in Table 1 with their range in search space, dimension and optimal value. These functions have two categories:

- (i) Group I: Unimodal minimization functions such as f_1, f_2, f_3, f_4, f_9 , and f_{10} . These functions are used for fast convergence.
- (ii) Group II: Multimodal minimization functions such as f_5, f_6, f_7 , and f_8 . These functions have large number of local points and test the performance evaluation of the algorithm to find a global optimal solution by avoid trapping with in local optima.

Simulations are performed in C++ based platform QT creator under Microsoft Windows 7 basic operating system. In each run, 1000 number of maximum iterations is being used as termination criteria. The value of α is taken as 0.1 and value of γ is 1.0. α varies between. $\alpha \leftarrow [0, 1]$. The attractiveness parameter, i.e., β_0 and betamin are taken as 1.0 and 0.2.

In order to produce better results, mean value on each benchmark function is calculated by taking a mean of 20 Monte Carlo runs and the population size for all these three algorithms is set to 20 [23, 24].

7 Simulation and Results

The results clearly demonstrate that the proposed algorithm is superior in performance as shown in Table 2. The results reveal that proposed algorithm gives better accuracy and have better convergence than other two algorithms.

- (i) Group I: This proposed algorithm computes better results on five unimodal functions out of total six unimodal functions that are f_1, f_3, f_4, f_9 , and f_{10} . Firefly Algorithm gives better result on f_2 function.
- (ii) Group II: The proposed algorithm perform well on two multimodal functions that are: f_6 and f_7 and Firefly Algorithm is better on another two multimodal functions that are: f_7 and f_8 . Flower Pollination Algorithm is efficient and has computed results but not as much as efficient in comparison of other two algorithms.

Table 1 Various standard benchmark functions with their formulas

Benchmark functions	Formula	Dimension	Range	Optima
Sphere	$f_1(x) = \sum_{i=1}^n x_i^2$	30	(-100, 100)	0
Alpine	$f_2(x) = \sum_{i=0}^n x_i \sin x_i + 0.1x_i $	30	(-10, 10)	0
Cigar	$f_3(x) = x_1^2 + \sum_{i=2}^n x_i^2$	30	(-10, 10)	0
Step	$f_4(x) = \sum_{i=0}^{n-1} (x_i + 0.5)^2$	30	(-100, 100)	0
Ackley	$f_5(x) = -20 \cdot \exp\left(\frac{0.2}{\sqrt{n}} \sum_{i=1}^n x_i^2\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	30	(-32, 32)	0
Schaffer	$f_6(x) = (x_0^2 + x_1^2)^{\frac{1}{4}} + (50(x_0^2 + x_1^2)^{0.1} + 1)$	2	(-100, 100)	0
Rastrigin	$f_7(x) = (x_1^2 - 10 \cos 2\pi x_1 + 10)$	30	(-5.12, 5.12)	0
Rosenbrock	$f_8(x) = 100(x_1 - x_0^2)^2 + (1.0 - x_0)^2$	30	(-10, 10)	0
Bochachesky	$f_9(x) = x_0^2 + 2.0x_1^2 - 0.3 \cos(3\pi x_0) \cos(4\pi x_1) + 0.7$	2	(-100, 100)	0
Leon	$f_{10}(x) = 100(x_1 - x_0^3)^2 + (x_0 - 1.0)^2$	20	(-10, 10)	0

Table 2 Results of proposed algorithm in comparison to FA and FPA

Benchmark functions	FA/FPA algorithm	Firefly algorithm	Flower pollination algorithm
Sphere	3.80E-06	4.73E-05	3.41E+03
Alpine	3.46E-01	3.15E-01	1.43E+01
Cigar	8.35E-08	2.28E-06	5.10E+01
Step	7.5110+00	7.5455+00	4.35E+03
Ackley	9.66E-01	9.01E-01	1.32E+01
Schaffer	1.15E-03	3.63E-03	2.52E-02
Rastrigin	3.1796E+01	3.4525E+01	1.783E+02
Rosenbrock	5.26E-03	1.66E-03	1.70E-02
Bochachesky	1.20E-09	7.82E-08	4.23E-01
Leon	1.26E-11	2.35E-10	7.25E-03

The overall comparison of algorithms shows that the proposed hybrid algorithm possesses good position than others because the proposed algorithm is less complex as compare to others and its convergence is also fast also it avoids trapping with in local optima. The mean value is given in bold font in Table 2 indicates the superior value on a particular algorithm.

8 Conclusion

In this research paper, an FA/FPA algorithm is proposed that is based upon hybridization of FA and FPA. Simulations and experiments demonstrated the potential of new proposed algorithm. The biological processes of both these algorithms are combined to outcome an improved form of both these algorithms. This proposed algorithm removes some shortcomings of FA and FPA. In case of FA, the chances of trapping with in local optima are still there, and the solutions are changing as the optima approaches. So, to enhance the quality of solutions, it is important to reduce the effect of randomization [4]. In case of FPA, the convergence rate is not efficient as it takes a large amount of time to explore the search space effectively. But the new proposed algorithm has proved its efficiency in this case by using the concept of exploration and exploitation effectively. This proposed algorithm is advantageous in case of fast convergence and not to get trap with in local optima and it avoids premature convergence also. All these features contribute to good performance of this algorithm. Furthermore, it is possible to extend this proposed algorithm considering some kind of modifications and in combination with other algorithms to deal with real-world problems further effectively.

References

1. X. S. Yang, "Nature-Inspired Metaheuristic Algorithms", Luniver Press (2008).
2. M. Dorigo, and M. Birattari. "Ant colony optimization." Springer US, (2010), pp. 36–39.
3. James Kennedy, "Particle swarm optimization." Encyclopedia of Machine Learning. Springer US (2010), pp. 760–766.
4. X.S Yang, "Firefly algorithms for multimodal optimization." Stochastic algorithms: foundations and applications. Springer Berlin Heidelberg (2009), pp. 169–178.
5. X.S Yang, M. Karamanoglu, and X. He. "Flower pollination algorithm: a novel approach for multiobjective optimization." Engineering Optimization 46.9 (2014), pp. 1222–1237.
6. V. Gazi and K.M. Passino, "Stability analysis of social foraging swarms," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, Vol. 34, No. 1, (2004), pp. 539–557.
7. S. Arora and S. Singh, "Performance Research on Firefly Optimization Algorithm with Mutation." International Conference, Computing & Systems (2014).
8. S. Arora and S. Singh, "The firefly optimization algorithm: convergence analysis and parameter selection." International Journal of Computer Applications 69.3 (2013), pp. 48–52.
9. Rashedi, Esmat, H.N. Pour, and S. Saryazdi. "GSA: a gravitational search algorithm." Information sciences 179.13 (2009), pp. 2232–2248.
10. X.S Yang, "Flower pollination algorithm for global optimization." Unconventional Computation and Natural Computation. Springer Berlin Heidelberg (2012), pp. 240–249.
11. Elbeltagi, Emad, T. Hegazy, and D. Grierson, "Comparison among five evolutionary-based optimization algorithms." Advanced engineering informatics 19.1 (2005), pp. 43–53.
12. D. Karaboga and B. Basturk. "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm." Journal of global optimization 39.3 (2007), pp. 459–471.
13. D. Karaboga and B. Basturk. "On the performance of artificial bee colony (ABC) algorithm." Applied soft computing 8.1 (2008), pp. 687–697.
14. S. Arora, S. Singh, "A conceptual comparison of Firefly algorithm, Bat algorithm and Cuckoo search" International Conference on Computing, Communication and Networking Technologies IEEE (2013).
15. Fister, Iztok, X.S Yang, and J. Brest. "A comprehensive review of firefly algorithms." Swarm and Evolutionary Computation 13 (2013), pp. 34–46.
16. X.S. Yang, "Firefly Algorithm, Lévy Flights and Global Optimization", Research and Development in Intelligent Systems XXVI (Eds M. Bramer, R. Ellis, M. Petridis), Springer (2010), pp. 209–218.
17. X.S Yang, "Multiobjective firefly algorithm for continuous optimization." Engineering with Computers 29.2 (2013), pp. 175–184.
18. X.S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," International Journal of Bio-Inspired Computation, Vol. 2, No. 2 (2010), pp. 78–84.
19. B. J. Glover, Understanding Flowers and Flowering: An Integrated Approach, Oxford University Press (2007).
20. X. S Yang, Mehmet Karamanoglu, and Xingshi He. "Multi-objective flower algorithm for optimization." Procedia Computer Science 18 (2013), pp. 861–868.
21. X.S Yang, "Flower pollination algorithm for global optimization." Unconventional Computation and Natural Computation. Springer Berlin Heidelberg (2012), pp. 240–249.
22. Gandomi, A. Hossein, and A.H Alavi. "Krill herd: a new bio-inspired optimization algorithm." Communications in Nonlinear Science and Numerical Simulation 17.12 (2012), pp. 4831–4845.
23. Waser, M. Nickolas, "Flower constancy: definition, cause, and measurement." American Naturalist, 1986, pp. 593–603.
24. X.S Yang, M. Karamanoglu, and X. He. "Flower pollination algorithm: a novel approach for multiobjective optimization." Engineering Optimization 46.9 (2014), pp. 1222–1237.

Uniqueness in User Behavior While Using the Web

Saniya Zahoor, Mangesh Bedekar, Vinod Mane
and Varad Vishwarupe

Abstract In recent years, the amount of online content has grown in enormous proportions. Users try to collect valuable information about contents in order to find their way to relevant web pages. And a lot of research is going on to collect valuable service usage data and process it using different methods to know their behaviors. Many systems and approaches have been proposed in the literature which tries to get information about the user's interests by profiling the user. The objective of the paper is to profile users on their specific devices and the web usage patterns based on the keyboard and mouse usage, time spent on the web. By analyzing the usage patterns of various users, we prove that the patterns exhibited by any one user are different from other users.

Keywords User behavior · Web · User profiling · User modeling · Usage analysis

1 Introduction

It is human nature to be curious, to learn new things, to want to find out more. With the rise of the web, the urge to learn more has increased by leaps and bounds. The World Wide Web, WWW as it is better known, has become the single largest source of information to mankind and is always accessible to users at the fingertips. As user's level of engagement in using the web increases the volume of pages

Saniya Zahoor (✉) · Mangesh Bedekar · Vinod Mane
MAEER's MIT Pune, Kothrud, Pune, India
e-mail: saniyazmalik@yahoo.com

Mangesh Bedekar
e-mail: mangesh.bedekar@gmail.com

Vinod Mane
e-mail: vinod2789@gmail.com

Varad Vishwarupe
MIT College of Engineering, Kothrud, Pune, India
e-mail: varad44@gmail.com

accessed also increases. The diversity in users using the web also makes it difficult to build any generic user assisting systems which will cater to all users. The web as it is designed handles all types of users and web sites uniformly.

Capturing information about users and their interest is one of the main functions of user profiling. Much of the research has been done on profiling in the field of recommender system and various profiling techniques have been evolved around the time. All these are done to know the behavior of the user so that we can understand his interests completely [1]. There are a lot of applications of the same like it can help to provide more relevant recommendations to the users and hence can help the e-commerce sites to increase their sales to even higher levels.

In this paper we propose means and mechanisms to capture, store, and analyze the user's entire web usage behavior as logs on the user's machine locally. This also serves the purpose of creating the user profile. This profile then can be used to make the web more personal to the user. The main objective of this paper is to analyze the behavior of different users on their respective devices and hence prove that the taste of each user is different when it comes to the way they access the web and search the web pages.

There are a lot of factors on which we have focused on and these are:

- Number of URLs visited by different users in a day.
- Time spent by different users in a day.
- Keyboard activities performed by different users in a day.
- Mouse activities performed by different users in a day.
- Time spent for the same URLs by different users.
- Order in which same URLs are opened by different users.
- Order in which same URLs are closed by different users.

The analysis of the user behavior is divided into two phases:

- In the first phase, two users work on two different systems and it was seen that they differ in many aspects like the number of URLs visited, time they spent on the URLs, keyboard, and mouse activities.
- In the second phase, two users are provided with the same set of URLs to visit. It was seen that the two users still have unique behavioral patterns with respect to, the manner in which they open and close URLs, the spent time on each URLs, the amount and type of keyboard and mouse activities performed on the URLs. All this data indicates variation in user behavior which can be easily identified, quantified, and later analyzed.

2 Related Work

References [2–7] have analyzed the behavior of the user from the time spent on the browser. References [3, 7–13] worked on mouse activities to analyze the behavior of the user. References [14, 15] have considered bookmark, print, time spent, and

scroll activities to understand user’s interest and then analyze his behavior. References [4, 16, 17] have focused on text selection in web pages as an implicit interest indicator of the user.

The existing systems studied, tend to characterize the user based on his web access patterns. It does seldom compare individual user access patterns with other users on same device or different devices so as to identify uniqueness of usage patterns and to what extent if any.

3 Proposed System

The implementation starts with installation of XAMPP (Cross, Apache, MySQL, PHP, PERL) server on the client’s machine and addition of number of database tables that stores various actions that user performs on the web page like time spent by a user on web pages, various mouse and keyboard patterns that a user makes while his usage on the web page and so on.

Greasemonkey is a Mozilla Firefox extension that allows users to install scripts that make on the fly changes to web page content after or before the page is loaded in the browser. The changes made to the web pages are executed every time the page is viewed, making them effectively permanent for the user running the script. With the help of Greasemonkey, we install out scripts on the browser so that we were able to capture all the parameters needed for the methods implementation.

JavaScript (JS) is an interpreted computer programming language. JavaScript along with tools of PHP and AJAX are used for capturing the data which directs it to the database where it is stored. The XAMPP server is where the database resides on the user’s machine locally. The JavaScript directs the data captured to the PHP page which stores the data in the corresponding table of the database in the XAMPP server. The system is implemented by writing JavaScripts for time, keyboard and mouse activities which are as listed in Table 1.

The user performs his day-to-day web tasks transparently in the system. There is no change required by the user so as to capture his actions. After the entire user-initiated events are captured, all these data along with the URL of the web page are stored into the database table as illustrated in Fig. 1. From this, the users’ behavior on the web page can be analyzed well.

Table 1 Attributes considered for user behavior analysis

Time	Keyboard	Mouse
Time the URL is opened, time the URL is closed, time spent on each URL	Copy, paste, save, print, bookmark, search, highlight	Scrolls, clicks

url	SaveAs	copy	paste	print	AddBookmark	clicks	times	highlight	longscroll	shortscroll
http://localhost/enter/arraytest.php	0	0	0	0	0	0	3	0	0	0
http://localhost/examples	1	2	2	2	1	11	204	2	95	200
http://localhost/phpmyadmin/#	0	0	0	0	0	5	0	0	0	0
http://localhost/phpmyadmin/#PMAURL-0:index.php?db...	0	0	0	0	0	3	0	0	0	0
http://localhost/phpmyadmin/#PMAURL-0:index.php?db...	1	1	1	1	0	19	181	0	131	265
http://localhost/security/index.php	0	0	0	0	0	0	36	0	0	0
http://localhost/security/security.php	1	1	0	1	0	15	0	2	0	0
http://localhost/xampp/	5	0	0	0	0	0	1705	0	0	0
http://localhost/xampp/blorhythm.php	0	2	0	0	1	11	71	0	26	43
http://localhost/xampp/guestbook-en.pl	3	1	1	1	0	6	0	4	0	0
http://localhost/xampp/head.php	1	2	1	1	0	8	1638	1	0	0
http://localhost/xampp/navi.php	1	1	1	1	0	36	0	0	0	0
http://localhost/xampp/phonebook.php	0	1	1	0	0	9	0	2	0	0
http://localhost/xampp/phpinfo.php	3	1	2	2	1	29	193	4	76	262
http://localhost/xampp/start.php	3	6	4	3	1	27	0	0	168	445

Fig. 1 Database snapshot

4 Result Analysis

The experimentation was instrumented as mentioned in Sect. 3 and tested. The user behavior was elaborately tested and analyzed in different conditions. It was tested for a single user as well as multiple users. The behavior of users was tested for a single site and also for multiple sites. The users were asked to repeat tasks on various days to nullify the effect of bias if any. The task given to users was carried out in isolation to mimic the user’s usual web browsing pattern.

It was required to analyze the users’ behavior based on quantifiable values directly observable from the browser, like—time spent, keyboard usage and mouse activities performed, etc. It is clearly observed that the users behaviors differ on most fronts, viz., same sites, same day visits, multiple sites, number of keyboard and mouse movements, time spent on each sites, etc. It goes on to validate our claim that users online behavior in terms of time spent, keyboard usage, mouse movements is different from each other for same tasks, on same sites as well as on different sites.

A sample of the results obtained from the above-elaborated system is mentioned in the following sections.

4.1 Result Analysis on Different Sets of URLs for Two Users in One Day

A particular test analysis was done on two users for their usage on different sets of URLs on a single day. Figure 2 shows the time spent verses number of different sets of URLs for two users on a single day.

Fig. 2 Time spent verses number of different sets of URLs for two users

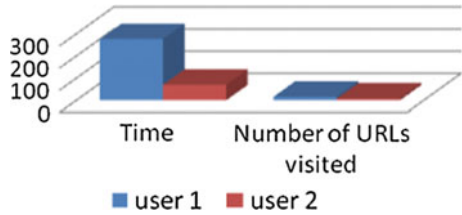


Fig. 3 Keyboard activities on different sets of URLs for two users

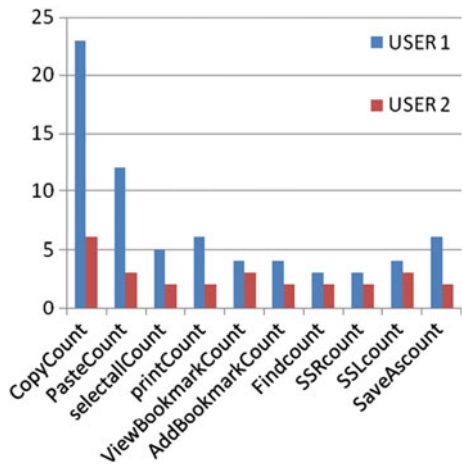


Figure 3 shows the keyboard activities on different sets of URLs for two users on a single day.

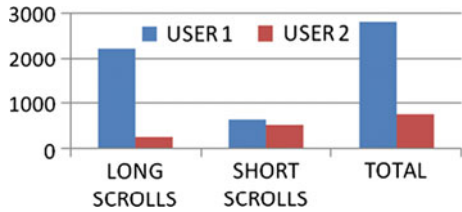
Figure 4 shows the mouse activities on different sets of URLs for two users on a single day.

Figure 5 shows the overall activities on different sets of URLs for two users on a single day.

4.2 Result Analysis on Same Set of URLs for Two Users in a Single Day

Another particular analysis was done on two users for their usage on same set of URLs on a single day. Figure 6 shows the opening time of same set of URLs by two users on a single day.

Fig. 4 Mouse activities on different sets of URLs for two users



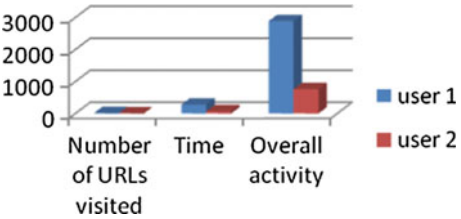


Fig. 5 Overall activities on different sets of URLs for two users

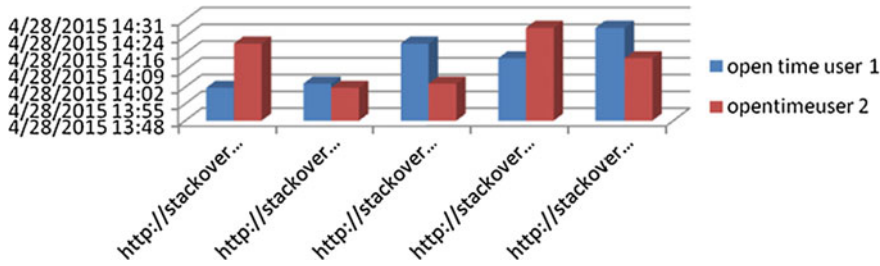


Fig. 6 Opening time of same set of URLs by two users

Figure 7 shows the closing time of same set of URLs by two users on a single day.

Figure 8 shows the time spent on same set of URLs by two users on a single day.

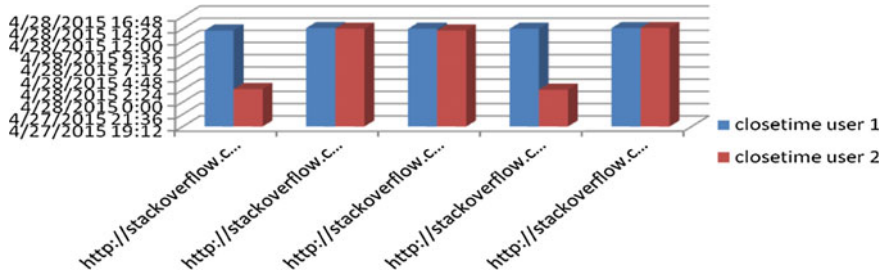
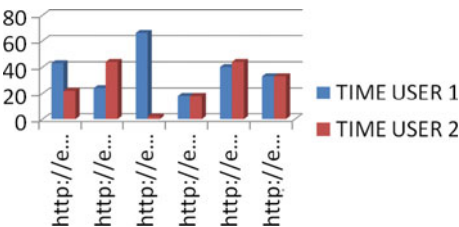


Fig. 7 Closing time of same set of URLs by two users

Fig. 8 Time spent on same set of URLs by two users



5 Conclusions

The user's online behavior can be analyzed very well from the navigation and usage behavior exhibited on the web page done in the browser. It is possible to capture almost all of the actions that the user does on the web page. If we consider the order of occurrence of the exhibited behavior and the granularity of the behavior, it will result in fine-grained understanding of user's interest for profiling the user.

Every user has a unique pattern of accessing the web, whether on any web page of the user's choice or if asked to visit a given set of web pages to satisfy any particular task. This paper focused on the way a user behavior is different compared to other users for completing the same task and the patterns of such unique characteristics can lead to actually differentiating one user from other users. This leads to having a unique user profile for every user. This user profile can be made richer and complete if the content accessed by the user is related to the browsing patterns exhibited by the user.

Once we understand the complete behavior of the user, this information can be used for and by applications like e-commerce sites and various recommendation systems to give focused services to each user differently.

References

1. Ed Chi H, Rosien A., Heer J, Jack L, Intelligent Discovery and Analysis of Web User Traffic Composition, WEBKDD 2002, Mining Web Data for Discovering Usage Patterns and Profiles, LNAI 2703, Lecture Notes in Computer Science, 2003, pp. 1–16.
2. Claypool M, Le P, Wased M, Brown D, Implicit interest indicators, Proceedings of the 6th international conference on Intelligent user interfaces, January 14–17, 2001, Santa Fe, New Mexico, USA. pp. 33–40.
3. Faucher J, McLoughlin B, Wunschel J, Implicit web user interest. Technical Report MQP-CEW-1101, Worcester Polytechnic Institute, Spring 2011.
4. Hauger D, Paramythis A, Weibelzahl S, Using Browser Interaction Data to Determine Page Reading Behavior, UMAP'11, Proceedings of the 19th International Conference on User modeling, adaption, and personalization, Girona, Spain, July 11–15, 2011, pp. 147–158.
5. Kellar M, Watters C, Duffy J, Shepherd M, Effect of task on time spent reading as an implicit measure of interest, Proceedings of the 67th Annual Meeting of the American Society for Information Science, 41(1), pp. 168–175.
6. Rastegari H, Shamsuddin S. M, Web Search Personalization Based on Browsing History by Artificial Immune System, International Journal of Advances in Soft Computing and Its Applications, Volume 2, Number 3, November 2010, ISSN Print: 2074–8523, ICSRS Publication.
7. Velayathan G, Yamada S, Can We Find Common Rules of Browsing Behavior, WWW 2007, 16th International World Wide Web Conference, Workshop on Query Log Analysis, May 8–12, 2007, Banff, Canada.
8. Dupret G, Lalmas M, Absence time and user engagement: evaluating ranking functions, Proceedings of the sixth ACM international conference on Web search and data mining, February 04–08, 2013, Rome, Italy, pp. 173–182.

9. Merlin B, Improving the expressiveness of navigation task to evaluate users' interest for pages, IHC2010, 2nd Workshop on Aspects of the Human-Computer Interaction for the Social Web, October 5–8, 2010, Belo Horizonte, Brazil, pp. 20–28.
10. Mueller F, Lockerd A, Cheese: tracking mouse movement activity on websites, a tool for user modeling, CHI '01, Extended Abstracts on Human Factors in Computing Systems, March 31–April 05, 2001, Seattle, Washington, pp. 279–280.
11. Shen X, Tan B, Zhai C. X, Context-sensitive information retrieval using implicit feedback, Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, August 15–19, 2005, Salvador, Brazil, pp. 43–50.
12. Sugiyama K, Hatano K, Yoshikawa M, Adaptive web search based on user profile constructed without any effort from users, WWW'04, Proceedings of the 13th international conference on World Wide Web, May 17–20, 2004, New York, NY, USA, pp. 675–684.
13. Luis A. Torres L, Hernando R. V, A Gesture Inference Methodology for User Evaluation based on Mouse Activity Tracking, IHCI 2008, Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction, Amsterdam, The Netherlands, July 25–27, 2008.
14. Holub M, Bielikova M, Estimation of user interest in visited web page, WWW '10, Proceedings of the 19th international conference on World Wide Web, April 26–30, 2010, Raleigh, North Carolina, USA, pp. 1111–1112.
15. Teevan J, Dumais S, Horvitz E, Personalizing search via automated analysis of interests and activities, SIGIR '05, Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, New York, NY, USA, pp. 449–456.
16. Hijikata Y, Implicit User Profiling for On Demand Relevance Feedback, IUI'04, Proceedings of the 9th international conference on intelligent user interfaces, January 13–16, 2004, Madeira, Funchal, Portugal, pp. 198–205.
17. White R. W, Buscher G, Text selections as implicit relevance feedback, SIGIR '12, Proceedings of the 35th International ACM SIGIR Conference on Research and development in information retrieval, August 12–16, 2012, Portland, USA, pp. 1151–1152.

Prediction of ERP Outcome Measurement and User Satisfaction Using Adaptive Neuro-Fuzzy Inference System and SVM Classifiers Approach

Pinky Kumawat, Geet Kalani and Naresh K. Kumawat

Abstract Nowadays, ERP (enterprise resources planning) system is one of the very crucial and costly projects in the field of information systems for business investment. We report a practical approach which applies both the fuzzy logic analytical model and an expert judgment method support vector machines (SVM) classifier to predict whether the ERP software implementation project succeeds or fail. Here we develop an ANFIS model and SVM model approach, where ANFIS method uses the concept of fuzzy logic to predict the key ERP outcome “user satisfaction” using causal factors during an implementation as predictors which gives the prediction result 5.0000 which is an accurate result in comparison to existing prediction techniques such as MLRA and ANN, where SVM is a binary classifier model which tells about the prediction of good and bad performances of ERP project by dividing the whole dataset into two classes by user-defined condition, where the values of user satisfaction below 5 sets the output value 0 and the user satisfaction value above 5 sets the result 1 which gives the successful prediction results. The main objective of this research is to give satisfaction with user for better prediction results of ERP implementation success.

Keywords Artificial intelligence • ERP systems • Fuzzy logic • ANFIS • SVM • User satisfaction

Pinky Kumawat (✉) • Geet Kalani
Rajasthan College of Engineering for Women, Bhankrota, Ajmer Road,
Jaipur 302026, Rajasthan, India
e-mail: kumawat.pinky3@gmail.com

Geet Kalani
e-mail: kalanivmrgg@gmail.com

N.K. Kumawat
Department of Physics, Indian Institute of Technology Bombay, Powai,
Mumbai 400076, Maharashtra, India
e-mail: nareshkk@iitb.ac.in

1 Introduction

This enterprise resource planning (ERP) system is integrated and customized software-based system that solves many problems of system requirements in all the functional fields such as finance, sales, marketing, manufacturing, and human resources [1, 2]. Although expectations from the ERP systems are high, these systems have not always led to significant organizational improvements and most of the ERP projects become over the budget, late, and fail [3, 4]. Previous research depict that the failure of ERP projects is found due to the results of poor project communication, lack of the support of top management, existence of the cultural difference, low acceptance level of users and user dissatisfaction [5–10]. Although some researchers have discussed the flexibility measurement method of ERP system, however, the interaction and feedback relationships among criteria or indices are not taken into account in existing research results. Furthermore, the process of ERP outcome measurement has a good deal of uncertainty and vague information. Hence, the prediction techniques used for the satisfaction with user using ERP systems should be efficient to give best prediction results. In the present study the data onto the earlier research was used to develop predictive models for the implementation of ERP outcomes measured in terms of user satisfaction. So, the main motivation of this research is to predict the success or failure of ERP systems outcome measurements using the causal factors which are process, strategic, vendors and user satisfaction as outcome prediction result. Two prediction techniques, adaptive neuro-fuzzy inference system (ANFIS) and support vector machines (SVM) are used in this research work.

2 Literature Review

In this review of literature we depict various research views of ERP system prediction outcome measurements and user satisfaction. Botta Genoulaz et al. provided a survey to investigate the research activities related to ERP in recent days and found that the research on ERP systems has experienced an efficient development in recent years [11]. Various research models use many types of information systems, but not developed a model which is especially for ERP systems. Although, they provided basic general principles that could be useful for further research [12, 13, 14]. McAfee depicted the effect of ERP on the institution operational performance outcome [15]. Hence, the calibration of the enterprise resource planning (ERP) standard methods of the institutional procedures of the company has been considered a vital step in the procedure of exertion and acquires the attention on many scientists [16, 17]. Chien and Tsaur gave the model of DeLone and Mclean to describe the model's success in ERP systems and to identify the factors contributing to the high quality of ERP systems, the benefits of the use, and the individual performance [18]. In this manner, Ifinedo and Nahar get that the system quality and information quality are accepted as two main factors in the success rate and prediction of ERP systems [19]. Recently,

Chan et al. depicted a survey for good understanding of the approval for ERP systems in an individual manner [20]. This literature review of study provided a conceptual procedure to analyze the effect of the factors like compatibility, social impacts, the short-term consequences, and their impacts on the ERP use as outcomes. Sun et al. more recently studied the role of enterprise resources planning from several perspectives, namely the compatibility of work, identified usefulness, easily use, performance outcome measures and intended use on the performance of enterprise resources planning users and how these factors shape the use of ERP [21]. Perez-Bernal and Garcia-Sanche described that involvement of the user, training, and the managerial support are the tedious factors for ERP systems that connect directly to the users and customers depicting such infrastructures for implementing ERP systems [22]. In advance to these factors, Lo and Ramayah reported the effect of shared beliefs on the advantages of ERP within various users, containing engineers and managers [23]. The performance gives that, in an ERP system environment, satisfaction is significantly related to the performance factors.

3 Methodology

Following are the two methods used for ERP prediction by us:

1. Adaptive neuro-fuzzy inference system (ANFIS).
2. Support vector machines (SVM).

3.1 *Adaptive Neuro-Fuzzy Inference System (ANFIS)*

ANFIS is a class of adaptive networks which is equivalent to function of fuzzy inference system. Adaptive neuro-fuzzy inference system represents Sugeno e Tsukamoto fuzzy models. It uses hybrid learning algorithms using artificial neural networks and fuzzy reasoning, that is, ANN and FIS [24]. ANFIS system consists of five layers, which performs various actions in ANFIS, given in Fig. 1a.

3.2 *Support Vector Machines (SVM)*

SVM classify the data by finding best hyperplane that separates all data points of one class of the data into the other class. The hyperplane for an SVM means the one with the largest margin of the two classes where margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. We can use SVM when the data has two classes. Support vectors are the data points that are nearest to the separating hyperplane; these are the points that are on the boundary of the slab. Figure 1b illustrates the definition of SVM [25].

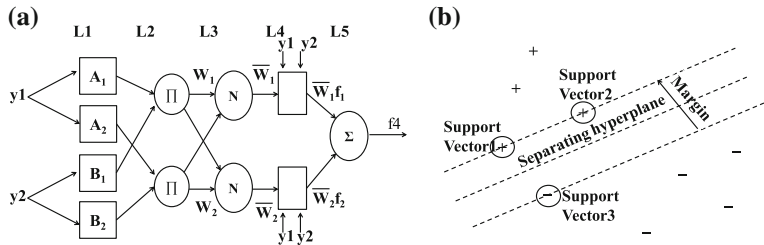


Fig. 1 **a** ANFIS 5 layer diagram with L1, L2, L3, L4, and L5 as layers: y_1 and y_2 used as input and z_1 as output function f_4 [26]. **b** SVM classifying the data by finding the best hyperplane

Table 1 Sample dataset/training and test dataset

S. no.	Process	Strategic	Vendor	User satisfaction
1	-1.38304	-0.77426	0.69434	4
2	-0.49719	-0.62034	2.14214	5
3	0.33652	-0.71615	0.50457	5
4	-0.56822	-0.08176	1.35287	5
5	-0.19034	-0.36148	0.69369	6
6	0.99333	-0.72094	-0.60319	6
7	-2.21601	-0.81793	-0.8998	1
8	1.12723	-0.32234	1.22514	6
9	0.86559	-1.52226	-0.70272	4
10	-0.28925	-1.53267	-1.49165	4

The sample dataset used in this paper and collected from prior research is given [27] in Table 1.

4 Modeling—Results and Discussions

4.1 Prediction Using ANFIS

ANFIS predicts the best results from other prediction techniques like MLRA and ANN. But it gives less efficient results than KNN classifiers. Here the prediction results from ERP implementation for the satisfaction with user as a failure or success of ERP project shows the various prediction results representation. We divide the data onto the train and test data onto prediction using ANFIS. Training data is shown in Fig. 2a.

The prediction result given by Eq. (1) is 5.0000 which is the accurate result given in dataset Table 1 as user satisfaction.

$$\text{Output} = \text{evalfis}([0.33652, -0.71615, 0.50457]', xyz) \quad (1)$$

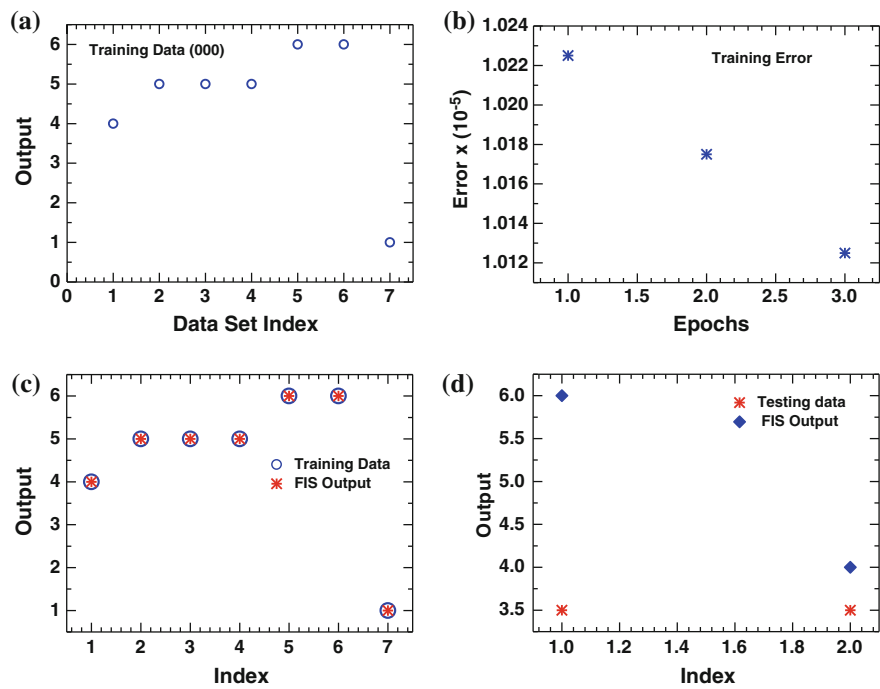


Fig. 2 **a** Training data used for ERP success prediction outcome gives the representation of training data where y axis shows the output that indicates the user satisfaction values corresponding the horizontal axis which shows dataset indices that indicates the row from which that input data values was obtained and the training data appears as o in the graph plot. With epoch = 3; error = 1.0125e-05. **b** Training error occurs to the ANFIS prediction using training data onto average testing error = 7.4111e-06, where, epoch means one iteration through the process of providing the network with an input and updating the network's weights, typically many epochs are required to train the neural network and training error is the difference between the training data output value and the output of fuzzy inference system corresponding to the same training data input value. In our training dataset three errors occur which are 1.022×10^{-5} , 1.017×10^{-5} and 1.012×10^{-5} corresponding to epochs 1, 2 and 3. Less training error represents efficient training in dataset in ANFIS and efficient training results indicates efficient testing of dataset for our ERP prediction. **c** Generation of fuzzy inference system with training data onto ANFIS with an index on horizontal axes which represents the 7 dataset taking for training the dataset and output on vertical axes which shows the user satisfaction corresponding output values in the given dataset. Hence it gives the training dataset presentation with FIS output of training dataset which shows good FIS generation corresponding the training data onto ANFIS and pretend to give good prediction in our project with average testing error = 1.8028. **d** Generation of fuzzy inference system with testing data onto ANFIS with an index on horizontal axes which shows the dataset indexes corresponding to the vertical axis that is the output user satisfaction values which we want to achieve. This graph show more difference between the testing data and FIS output which affects our result. Hence less the difference then the error is less and more the difference then the error is more

4.2 Prediction with SVM

This study tells that SVM classifies the data into two classes and find the best hyperplane which separates these classes. Hence, the SVM classifies here all the data into two classes according to user-defined condition. The condition applying here for dataset classification is the outcome of user satisfaction with less than the value 5 given class 1 of value 0 for all and the outcome of user satisfaction with the value above 5 given value 1 for class 2 for all corresponding outcome values. The SVM structure data used by this method is shown using **svmStruct** command which is given below.

```
>> svmStruct
svmStruct =
    SupportVectors: [7×3 double]
        Alpha: [7×1 double]
        Bias: -0.0143
    KernelFunction: @linear_kernel
    KernelFunctionArgs: {}
        GroupNames: [10×1 double]
    SupportVectorIndices: [7×1 double]
        ScaleData: [1×1 struct]
    FigureHandles: []
```

The given prediction input values from dataset are taking as input for SVM classification using following **xnew** command in MATLAB.

```
>> xnew= [-0.1903 -0.3615 0.6937]
Xnew = -0.1903 -0.3615 0.6937
```

This value of **xnew** from the sample dataset corresponds to the user satisfaction value which is above 5; hence give prediction result 1 after SVM classification as outcome of ERP implementation for the prediction and the satisfaction of user using following command in MATLAB where **svmclassify** classifies the values in **xnew** variable which contains values from training dataset for prediction using the information in a support vector machine classifier structure **svmStruct** and plot the **xnew** variable values in figure created using the **show** plot property with the **svmtrain** function and this plot appears only when the data is two dimensional.

```
>>results=svmclassify(svmStruct,xnew,'showplot',false);
>>results
results=1
```

The values corresponding to the user satisfaction below 5 gives the result 0 after using SVM classification as a prediction technique (Fig. 3).

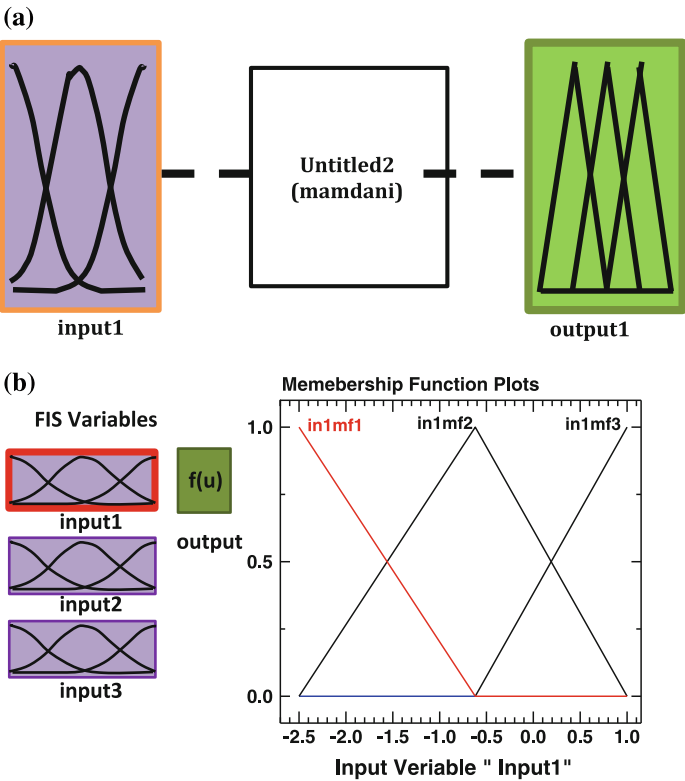


Fig. 3 **a** Mamdani model generation of fuzzy inference system with input1 in *left* and output1 in *right* and uses the technique of defuzzification of a fuzzy output. **b** Membership functions associated with all of the input and output variables for the entire fuzzy inference system. An element of the variable can be a member of the fuzzy set through a membership function that can take values of the range of 0–1. Membership function associated with the input and output parameters to be used in the FIS model can either be chosen by the user arbitrary based on the user’s experience or can also be designed using machine learning methods. **b** Represents all membership functions for the selected variable that is input1

5 Comparative Study

While comparing the proposed prediction techniques in this paper with existing ERP outcome measurement prediction techniques then we can conclude that ANFIS and SVM are efficient prediction techniques which give nearest accurate prediction results with user satisfaction outcome. The prediction results from same input values 0.33652, -0.71615, 0.50457 given by MLRA is 4.4676, given by ANN is 4.0706, and given by ANFIS is 5.0000 which shows that ANFIS gives best prediction output result whereas SVM classifies whole data into two classes and gives results according to user-defined condition. In this paper we define that user satisfaction values below 5 sets the output result 0 and user satisfaction values

above 5 sets the output result 1. In this paper the prediction results, satisfy the user-defined condition and gives output results that also satisfy the user for ERP implementation.

6 Conclusion and Future Scope

This research work has modeled the ERP implementation procedure, using the sample dataset with input predictors as process, strategic, and vendor and user satisfaction as outcome of prediction that impacts the success or failure of an ERP implementation. We develop ANFIS and SVM classifiers as prediction techniques. In these prediction techniques ANFIS gives best prediction results of ERP implementation and SVM also classifies the whole data into two classes for prediction and gives efficient result according to the user-defined prediction condition. With respect to this study we can further improve the prediction efficiency with respect to many other factors using other techniques which give the most accurate prediction results of ERP success or failure. Hence, the future work of ERP implementation research is that to improve the success of ERP projects by evolving best prediction techniques.

References

1. E.E. Watson, H. Schneider, "Using ERP in Education," *Communications of the AIS*, 1 (1998).
2. N. Basoglu, T. Daim, and O. Kerimoglu, "Organizational adoption of enterprise resource planning systems: a conceptual framework," *The Journal of High Technology Management Research*, in press (2007).
3. C. Soh, S. Kien, & J. Tay-Yap, "Cultural fits and misfits: is ERP a universal solution?" *Communication of the ACM*, vol. 43, no. 4, pp. 47–51 (2000).
4. V.B. Genoulaz, P.A. Millet, "An investigation into the use of ERP systems in the service sector," *International Journal of Production Economics*, vol. 99, pp. 202–221 (2006).
5. T.M. Somers, K.G. Nelson, "A taxonomy of players and activities across the ERP project life cycle," *Information and Management*, vol. 41, pp. 257–278 (2004).
6. M. Al-Mashari, A. Al-Mudimigh, and M. Zairi, "Enterprise resource planning: A taxonomy of critical factors," *European Journal of Operational Research*, vol. 146, pp. 352–364 (2003).
7. Y. Yusuf, A. Gunasekaran, and M.S. Abthorpe, "Enterprise information systems project implementation: A case study of ERP in Rolls-Royce," *International Journal of Production Economics*, vol. 87, pp. 251–266 (2004).
8. K.A. Gyampah, User involvement, ease of use, perceived usefulness and behavioral intention: A test of the enhanced technology acceptance model in an ERP implementation environment. *Proceedings of the 1999 Decision Sciences Annual Meeting*, New Orleans, LA, Nov. 20–23, pp. 805–807 (1999).
9. K.A. Gyampah, "Perceived usefulness, user involvement and behavioral intention: an empirical study of ERP implementation," *Computers in Human Behavior*, vol. 23, pp. 1232–1248 (2007).

10. F. Calisir & F. Calisir, "The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems," *Computers in Human Behavior*, vol. 20, pp. 505–515 (2004).
11. Botta-Genoulaz, V., Millet, P., and Grobot, B. "A survey on the recent research literature on ERP Systems", *Computers in Industry*, vol. 56(6), pp. 510–522, (2005).
12. Somers, T., and Nelson, K. "The impact of critical success factors across the stages of enterprise resource planning implementations", Paper presented at the Proceedings of the 34th Hawaii International Conference on System Sciences, January (2001).
13. Bradford, M., and Florin, J. "Examining the role of innovation diffusion factors on the implementation success of enterprise resource planning Systems", *International Journal of Accounting Information Systems*, vol. 4(3), pp. 205–225 (2003).
14. Hong, K., and Kim, Y. "The critical success factors for ERP implementation: an organizational fit perspective", *Information & Management*, vol. 40, pp. 25–40 (2002).
15. McAfee, A. "The impact of Enterprise technology adoption on operational performance: an empirical investigation", *Production and Operations Management*, vol. 77(1), pp. 33–53 (2002).
16. Chiplunkar, C, Deshmukh, S., and Chattopadhyay, R. "Application of principles of event related open Systems to business process reengineering", *Computers & Industrial Engineering*, vol. 45, pp. 347–374 (2003).
17. Van der Aalst, W. M. P., and Weijters, A. J. M. M. "Process mining: a research agenda", *Computers In Industry*, vol. 53(3), pp. 231–244 (2004).
18. Chien, S., and Tsaur, S. "Investigating the success of ERP Systems: Case studies in three Taiwanese high-tech industries", *Computers in Industry*, vol. 58(8), pp. 783–793 (2007).
19. Ifinedo, P., Rapp, B., Ifinedo, A., and Sundberg, K. "Relationships among ERP post implementation success constructs: An analysis at the organizational level", *Computers in Human Behavior*, vol. 26(5), pp. 1136–1148 (2010).
20. Chan, H., Siau, K., and Wei, K. "The Effect of Data Model, System and Task Characteristics on User Query Performance -An Empirical Study", the DATA BASE for Advances in Information Systems vol. 29 (1), pp. 31–49 (1998).
21. Sun, Y., Bhattacharjee, A., and Ma, Q. "Extending technology usage to work settings: The role of perceived work compatibility in ERP implementation. *Information & Management*, vol. 46, pp. 351–356 (2009).
22. Garcia-Sanchez, N., and Perez-Bernal, L. "Determination of Critical Success Factors in Implementing an ERP System: A Field Study in Mexican Enterprises", *Information Technology for Development*, 73(3), pp. 293–309 (2007).
23. Ramayah, T., and Lo, M. "Impact of shared beliefs on "perceived usefulness" and "ease of use" in the implementation of an enterprise resource planning system", *Management Research News*, 30(6), pp. 420–431 (2007).
24. J. S. R. Jang, "Adaptive-Network-Based Fuzzy Inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No. 3, pp. 665–685 (1993).
25. Support Vector Machines, <http://in.mathworks.com/help/stats/support-vector-machines-svm.html>.
26. Dr. Rajeev Gupta, Sarita Chouhan, Neha Bhuria "Comparative study of Institute based ERP based on ANFIS, ANN and MLRA", *International Journal of Engineering and Technical Research (IJETR)*, ISSN: 2321-0869, Volume-2, Issue-5, May (2014).
27. Chengaleth Venugopal, Siva Prasanna Devi, Kavuri Suryaprakasa Rao, "Predicting ERP User Satisfaction—an Adaptive Neuro Fuzzy Inference System (ANFIS) Approach", *Intelligent Information Management*, 2, 422–430 (2010).

UWB Antenna with Band Rejection for WLAN/WIMAX Band

Monika Kunwal, Gaurav Bharadwaj and Kiran Aseri

Abstract Owing to the progress in the field of wireless communication, UWB antenna has procuring more attention. A novel and compact ultrawideband antenna has been proposed for triple band notched rejection. The first rejection band is obtained by etching the C-shaped type slot in the partial ground structure. The second and last rejection bands are obtained by inserting the inverting and noninverting C-shaped type slots in the patch, respectively. The proposed antenna shows better radiation pattern, constant gain over the ultrawideband but not in the rejected frequency band. CST Microwave Studio Software is used for optimizing the parameters of UWB antenna having the capability for rejecting the three-stop band.

Keywords Band reject antenna • Inverted and noninverted C-shaped slots

1 Introduction

Nowadays, world has been moving toward augmented data rate and performance of antenna, UWB has been adapted due to its higher data rate over the large bandwidth. In 2002, Federal Communication Commission permitted frequency band ranging from 3.1 to 10.6 GHz for ultrawideband application [1]. Because of wide impedance bandwidth, augmented data rate, little power emission, compact in size, low profile, inexpensive, little power consumption, high radiation efficiency, and easy to fabricate on printed circuit board, ultrawideband antenna has gained an impetus and more attraction [2].

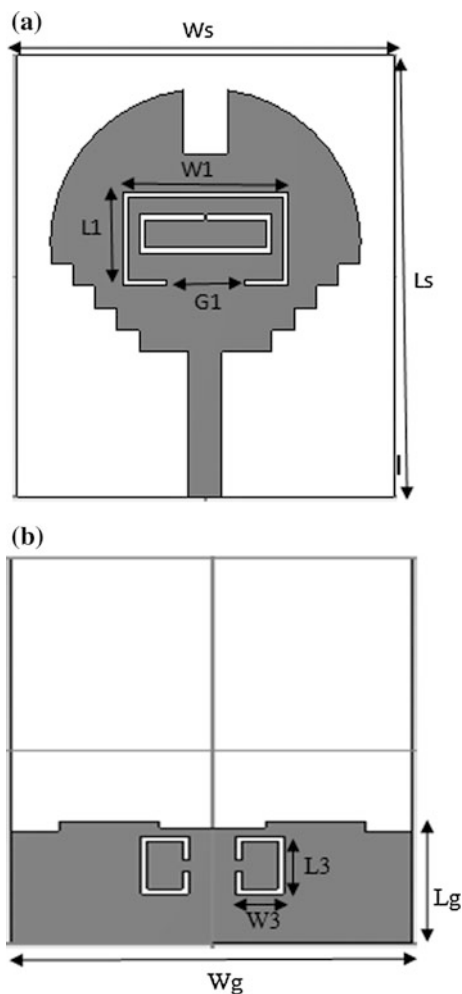
Monika Kunwal (✉) · Gaurav Bharadwaj · Kiran Aseri
Government Women Engineering College Ajmer, Ajmer, India
e-mail: monikakunwal@gmail.com

Gaurav Bharadwaj
e-mail: gwecaexam@gmail.com

Kiran Aseri
e-mail: kiranrids18@gmail.com

Fig. 1 Configuration of the desired antenna structure.

a Top view **b** Bottom view



Several planar antennas with ultrawideband characteristics have been reported in the literature such as triangular, square rectangular, circular, semicircular, annular, elliptical, and hexagonal, in shape [2–4]. Several single and multiband notched antennas have been reported in the literature [4–7]. Various methods are available to get band notch antenna to cut the slot not only on the patch but also on the ground or on the feed.

In this paper, an antenna having three-stop band has been proposed. C-shaped type slot is embedded in the ground for eliminating band from 5.1 to 6.03 GHz and for eliminating band from 2.45 to 2.74 GHz and 3.41–3.75 GHz, inverted C-shaped type slot and C-shaped type slot are introduced in the patch. For achieving the UWB, step slot is cut at the edges of the radiating patch. The required rejection band can be obtained by tuning the horizontal and vertical lengths of the desired band notched structure.

Table 1 Antenna parameter list

Parameters	Values (mm)
Ls	40
Ws	34
L1	8.5
W1	15
G1	7
L3	6
W3	4.05
Lg	12.3
Wg	34

2 Antenna Structure

Figure 1a and b shows configuration of desired antenna structure that has the capability of rejecting the three bands and it is located on x-y plane and z-axis is parallel to the normal direction [5]. Microstrip line is used as a feeding network for an antenna. The radiating patch is put on one side of FR-4 substrate and ground is placed on the backside of it. The thickness of FR-4 substrate is 1.6 mm with 4.4 dielectric constant and 0.02 loss tangent. For controlling the impedance bandwidth, the gap is introduced between the radiating patch and ground. The dimensions of antenna structure are optimized using software called CST software in order to achieve better impedance bandwidth and to get stable gain and radiation characteristics (Table 1).

3 Result and Discussions

The simulated VSWR is illustrated in Fig. 2. The simulated impedance bandwidth of the desired antenna is 2.34–10.5 GHz, for $S_{11} \leq -10$ dB. There are three stop bands in the frequency ranges 2.43–2.62 GHz, 3.35–3.69 GHz, and 4.94–5.96 GHz,

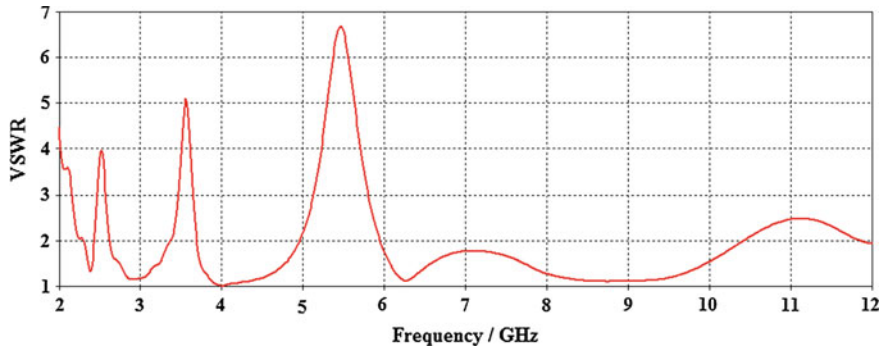
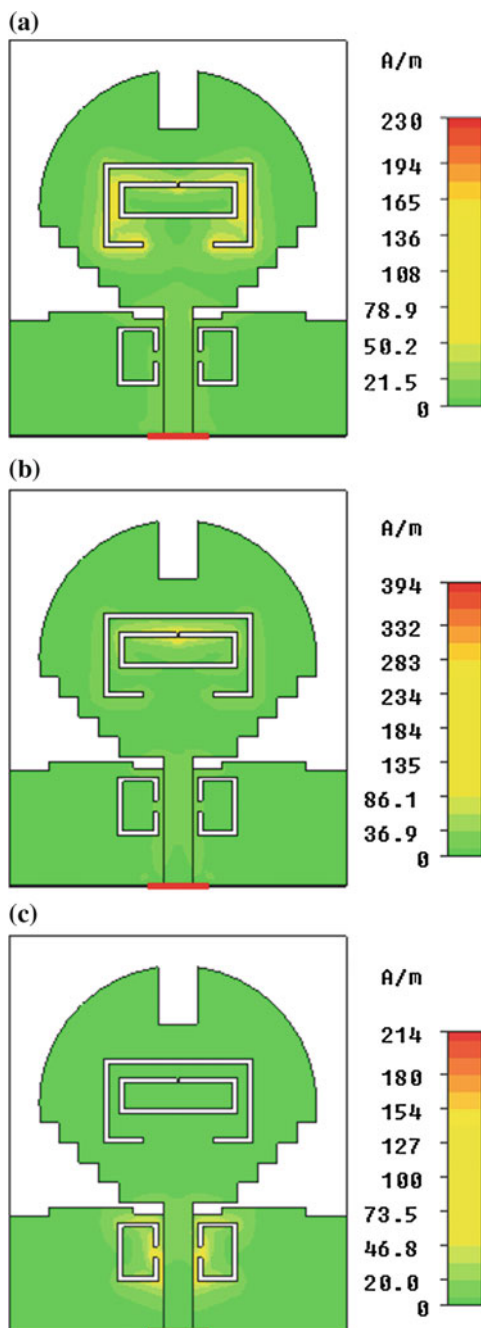


Fig. 2 The VSWR of desired antenna structure for rejecting three bands

Fig. 3 The surface current distribution of desired antenna structure at 2.55 GHz, 3.59 GHz and 5.5 GHz



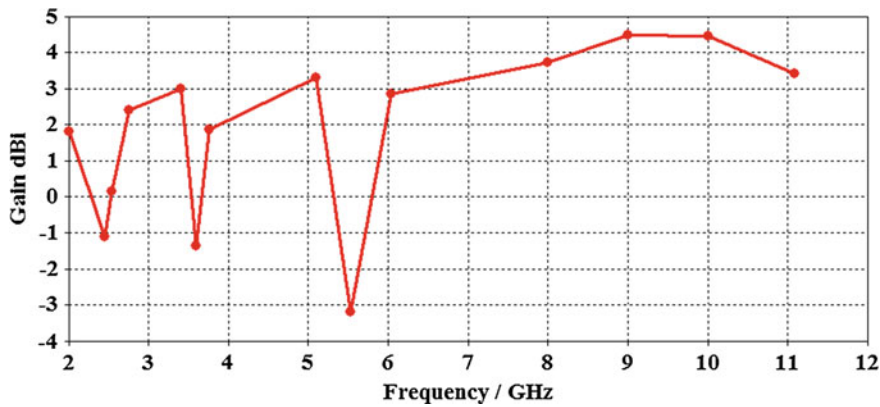


Fig. 4 The simulated gain of desired antenna structure

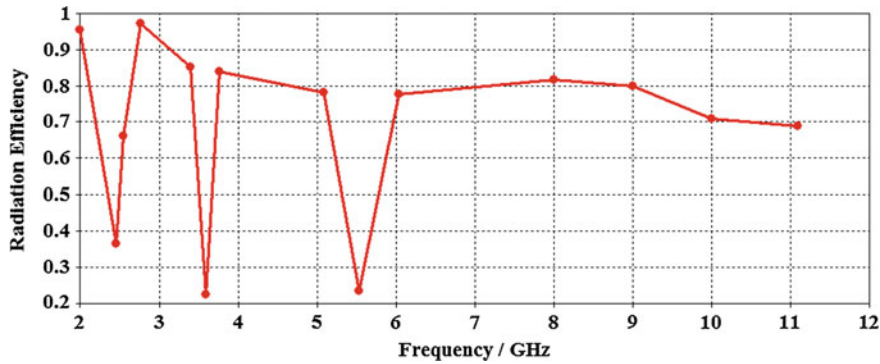


Fig. 5 The simulated radiation efficiency of desired antenna structure

for $VSWR > 2$. Therefore, these stop bands are used to avoid interference with 2.5 or 3.5 GHz WiMAX and 5.5 GHz WLAN band.

For understanding the mechanism of rejection characteristics, surface current distribution has been investigated at different rejection frequency band. The simulated surface current is generally concentrated around the inverted C-shaped type slot at the 2.55 GHz and there is a little current in the C-shaped type slot that placed on the ground.

Figure 3b, shows the surface current at 3.59 GHz generally concentrated at edges of C-shaped type slot. In Fig. 3c, the current is primarily concentrated on the C-shaped type slot that is embedded in the ground plane.

The simulated gain is shown in Fig. 4. Almost stable gain is obtained over the entire operating band. There is a sharp decrease of gain at 2.54, 3.59, and 5.54 GHz.

Simulated radiation efficiency is shown in Fig. 5. Radiation efficiency is almost 70–80 % throughout working band but not in the rejected bands. There is a sharp decrease of efficiency at the central notch frequency (Figs. 6, 7, 8).

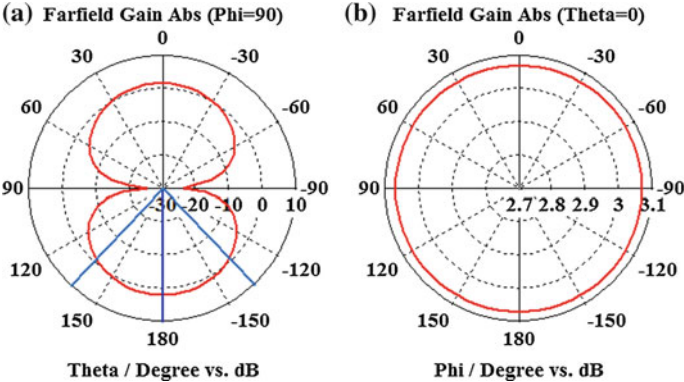


Fig. 6 The simulated radiation pattern of desired antenna at 2 GHz frequency. **a** E plane **b** H plane

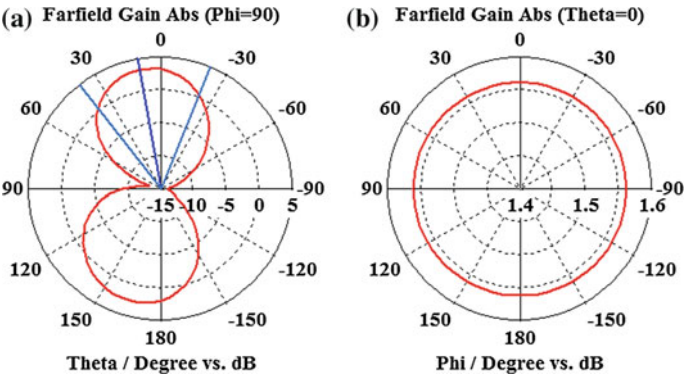


Fig. 7 The simulated radiation pattern of desired antenna at 5.09 GHz frequency. **a** E plane **b** H plane

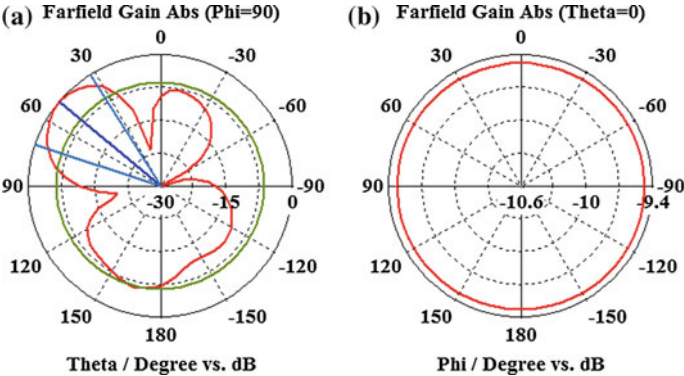


Fig. 8 The simulated radiation pattern of desired antenna at 10 GHz frequency. **a** E plane **b** H plane

4 Conclusion

Benefits of this antenna are easy to assemble, low cost and simple structure. The fundamental parameters of the antenna are return loss, radiation patterns, and bandwidth. All frameworks satisfy the acceptable antenna standard and the satisfactory results are observed. The three stop bands are attained by introducing the inverted and noninverted C-shaped type slots in the patch and C-shaped type slot in the ground. The UWB antenna with band notch features is expected to be good option to incorporate with portable devices.

References

1. G. P. Gao, M. K. Yang, S. F. Niu, and J. S. Zhang: Study of a Novel U-Shaped Monopole UWB Antenna by Transfer Function and Time Domain Characteristics, *Microwave and Optical Technology Letters*/vol. 54, no. 6, pp. 1532–1537 (2012).
2. Jianxin Liang, Choo C. Chiau, Xiaodong Chen, and Clive G. Parini: Study of a Printed Circular Disc Monopole Antenna for UWB Systems, *IEEE Transactions on Antennas And Propagation*, vol. 53, no. 11 (2005).
3. C.C Lin, Y.C. Kan, L.C. Kuo, and H.R. Chuang: A planar triangular monopole antenna for UWB communication, *IEEE Microwave Wireless Compon Lett* 15, pp. 624–626 (2005).
4. Jitendra Kumar Deegwal, Ashok Kumar, Sanjeev Yadav, Mahendra M. Sharma, and Mahesh C. Govil: Ultrawideband truncated rectangular monopole antenna with band-notched characteristics, *IEEE Symposium on wireless Technology and Applications* (2012).
5. Mohammad Nasser Moghadasi: A simple UWB microstrip fed planar rectangular slot monopole antenna, *Loughborough Antenna and Propagation Conference* (2011).
6. Sanjeev K. Mishra, Rajiv K. Gupta, Avinash Vaidya and Jayanta Mukherjee: Low-cost, compact printed circular monopole UWB antenna with 3.5/5.5 GHz band notch characteristics, *Microwave and Optical Technology Letters* vol. 54, pp. 242–246 (2012).
7. Chu, Qing-Xin and Ying-Ying Yang: A Compact Ultrawideband Antenna with 3.4/5.5 GHz Dual Band Notched characteristics, *IEEE Transactions on Antennas and Propagation*, vol. 56, issue, 12, pp. 3637– 3644 (2008).

Consequence of PAPR Reduction in OFDM System on Spectrum and Energy Efficiencies Using Modified PTS Algorithm

Luv Sharma and Shubhi Jain

Abstract The manuscript presents here establishing relation among energy efficiency, spectral efficiency and PAPR reduction. The results are shown by simulation over OFDM technique in MATLAB. It has been showed in results that with PAPR reduction, system can attain high SE and EE comparing with other system without PAPR reduction. Also we have evaluated the results over the clipping, filtering, PAPR reduction in OFDM, PTS4, and PTS8 algorithms. The desirable result can also be achieved using the higher order PTS algorithms. Also we have analyzed the relations with PAPR, spectrum efficiency, and energy efficiency of the OFDM Systems. The outcome using the PTS algorithm boosts the efficiency of the system by falling down the PAPR in OFDM systems.

Keywords Partial transmit sequence (PTS) • Energy efficiency (EE) • High power amplifier (HPA) • Orthogonal frequency division multiplexing (OFDM) • Peak-to-average power ratio (PAPR) • Spectrum efficiency (SE) • Input backoff (IBO)

1 Introduction

The need of high data rates in communication system is increasing very rapidly day by day. There is continuous demand for broadcast structure that can sustain such higher data rates for faster communication. On the contrary of the superior data rates there is also the requirement of the low power consumption in smart devices including smartphones, handheld computing devices, etc. It can also be said that if energy efficiency is achieved there exists a limitation of spectrum. Since spectrum resource is also very scarce and very low available for particular communication systems. Thus there is required a smart system that can comply both the energy

Luv Sharma (✉) • Shubhi Jain

Swami Keshvanand Institute of Technology, Gramothan & Management, Jaipur, India
e-mail: engineerluv@gmail.com

Shubhi Jain

e-mail: shubhijain19@gmail.com

efficient condition and spectrum demand to be satisfied. Thus the expected communication system required high data rate as well as highly efficient power usage. Amongst the various communication techniques for modulation and multiplexing, the OFDM system is treated to be the most efficient nowadays for achieving desirable SE and EE, multipath delay spread tolerance, power efficiency, and other important factors for a better communication system [1]. One can consider the OFDM system for the high data rates over microwave access techniques. The foremost constraint for the OFDM system is its high peak-to-average power ratio (PAPR) of broadcast signal.

Also in OFDM scheme result is superposition of several subcarriers, causing the increase in immediate power output than that of the mean power of the system. Thus, it requires HPA which is expensive and has very low efficiency. Also linearity of the system degraded resulting in distortion and degradation of SE and EE. To resolve the above problem the input backoff (IBO) of HPA must be larger than PAPR to avoid nonlinear distortion. If it is not so then this nonlinear distortion will result in reduction of the data rates. Also there exists a limitation that the power consumption increases with the increase in IBO. Now to improve the efficiency of HPA there is a major requirement of PAPR reduction resulting in power saving and thus improving the EE and SE performances of OFDM. SE and EE performances are already discussed in various literatures [2–4], so we are concerning only the method for the PAPR reduction to improve EE and SE performances. The nonlinear distortion noise is also related with the PAPR as discussed earlier so on reduction of the PAPR it will also reduce thus improving the data rates and reducing power consumption of overall communication systems [5]. There are several PAPR reduction techniques available like signal scrambling technique including selective mapping (SLM), partial transmitted sequence (PTS), interleaving technique, tone reservation (TR), tone injection (TI), and signal distortion technique includes peak windowing, envelope scaring, peak reduction carrier, clipping, and filtering [6, 7]. Amongst all the above schemes we are using the PTS scheme compared to other PAPR reduction techniques. PTS4 and PTS 8 schemes are for reducing PAPR and improving the SE and EE performances. The rest part of the paper includes review of the OFDM system, PAPR, SE, and EE, respectively. Nonlinearity and power consumption of HPA are discussed and finally the relation between PAPR reductions SE and EE derived for clipping, PTS 4, and PTS 8 schemes followed by conclusions [6].

2 About OFDM

OFDM is a generally pertinent wireless broadcast arrangement that involves tower capacity broadcast and high bit rate or information rates. The OFDM signal is the outcome of the composite signal produced by multiplexing the modulated signals [2, 4]. The conventional frequency division multiplexing (FDM) is prime basis for the technology and it is considered as advancement over this conventional

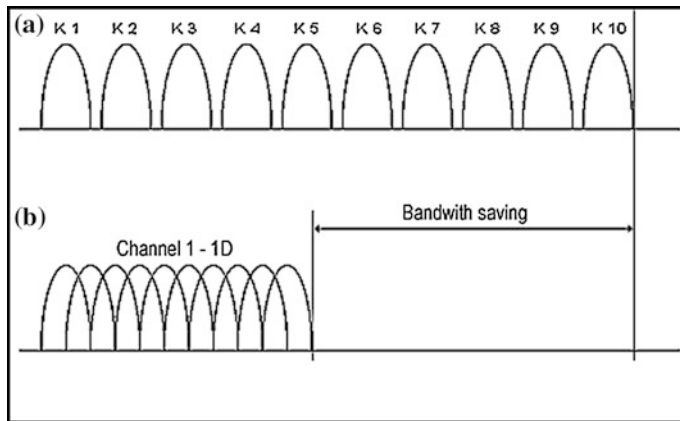


Fig. 1 OFDM spectrum

method which is adopted to hold only one signal over one conduit. The dispersive channel requires more advances and smooth transmission of signal and thus multicarrier modulation technique becomes a distinctive appearance for such transmission. The OFDM process includes a high rate data stream that is alienated into many low data streams. Further these streams are then multiplied by equivalent carrier frequency signals that are orthogonal to each other [8, 9]. The result is that the different carriers are orthogonal to each other, that is, they are absolutely autonomous of one another. This can also be achieved by placing the carrier exactly at the nulls in the modulation spectra of each other as shown in Fig. 1.

In OFDM transmission, the composite data representation slab $a = (a_0, a_1, \dots, a_{N-1})$ is conceded all the way through an N Point inverse fast Fourier transform (IFFT) accomplishing discrete time domain section to be broadcast.

The above statement concluded that the broadcast signal illustration is represented as $b_n^i = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} a_m^i e^{\frac{j2\pi m n}{M}}$

where i is the OFDM representation key, b_n^i is the data representation broadcast above m^{th} subcarrier.

3 About PAPR

PAPR of OFDM signals $x(t)$ is typified as the relation flanked by the highest instantaneous power and its average power [10–12]. The PAPR representation of the time sphere model progression $b = (b_0, b_1, \dots, b_{N-1})$ is defined as

$$\text{PAPR}(b) = \frac{\text{Max}|b_n|^2}{\frac{E\|b_n\|^2}{N}} = \frac{P_{\text{peak}}}{P_{\text{average}}}$$

where P_{peak} represents peak output power, P_{average} means average output power, $\|\cdot\|$ signify the norm of the enclosed vector b_n .

Since $b(n)$ is random, the PAPR is also a random variable. Therefore, complementary cumulative distribution function (CCDF) is for perpetuity that describes the statistical possessions of the PAPR in OFDM systems, i.e.,

$$\text{CCDF}_\lambda = \Pr\{\text{PAPR} > \lambda\} \quad [13-15],$$

where λ is a constant

The definition of the SE and EE in OFDM systems can be written as

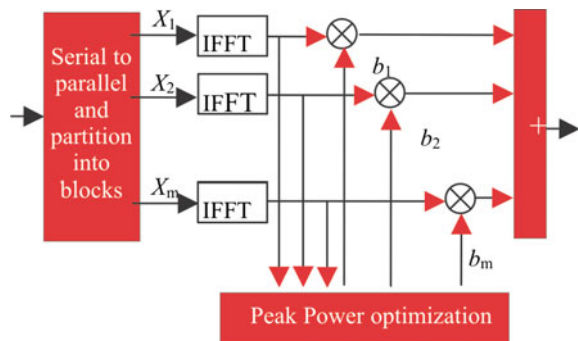
$$\eta_{\text{SE}} = \frac{R}{B}$$

$$\eta_{\text{EE}} = \frac{R}{P_{\text{hpa}}}$$

4 PTS (Partial Transmit Sequence)

Partial Transmit Sequence (PTS) The PTS technique involves the participation data block of N Symbols that are partitioned into disjoint sub blocks before the signals are transmitted. Some more issue that could authorize the PAPR reduction presentation in PTS are subblock partitioning, technique of the division of the subcarriers into multiple disjoint subblocks. There are three categories of subblock partitioning method: adjoining, interleaved, and pseudorandom partitioning. The most important advantage of this PTS technique is that it is compatible with an uninformed quantity of subcarriers and any modulation format (Fig. 2) [15–18].

Fig. 2 Flow diagram of the PTS technique showing working of various blocks [4]



The operating point of the HPA is set as IBO (input backoff) and defined as

$$\text{IBO} = 10 \log_{10} \frac{P_{\max}}{P_{\text{avg}}}$$

where $P_{\max}$ denoted the saturation input power of the HPA and P_{avg} is the average power of the input signals [19].

5 Performance Analysis

In this section, using MATLAB for simulation both theoretical analysis and numerical simulations demeanor are assessed by the effect of the PAPR reduction on EE and SE performances in OFDM systems [19]. A preassumption is considered that the circuits of the devices are unchanged and the over consumption of the other circuit devices P_c remains the same when the PAPR is reduced. Also another assumption is taken that OFDM signal is normalized as $T = 1$, and the bandwidth is $B = N$. For all simulations, quadrature phase shift keying (QPSK) employed sub-carriers $N = 64$ and SNR = 15 dB.

The phase rotation factor is $\{+1, -1\}$ and No. of subblocks $V = 4$ and $V = 8$, respectively. Results from figure PAPR of original OFDM signal is 12.45, CCDF = 10^{-4} ,

PAPR reduction by 3.13 and 5.05 dB at $V = 4$ and $V = 8$ is achieved (Figs. 3, 4, 5, 6 and 7).

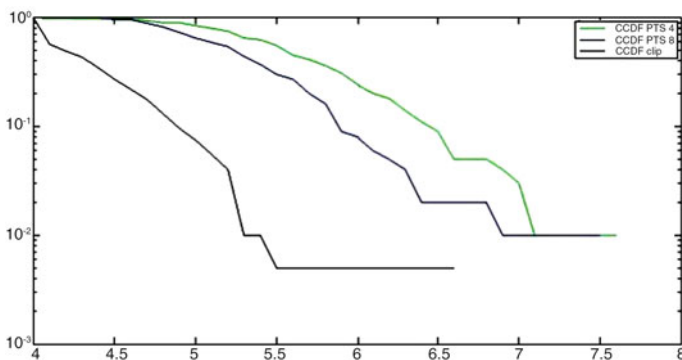


Fig. 3 PAPR lessening for the various PTS scheme taking $V = 4$ and $V = 8$

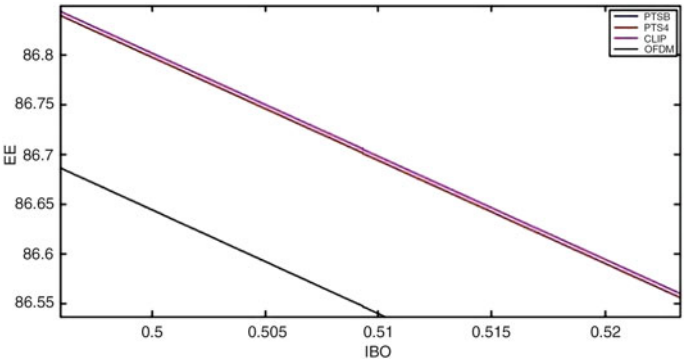


Fig. 4 Comparison of various signals for clip, PTS 4 and PTS8 techniques

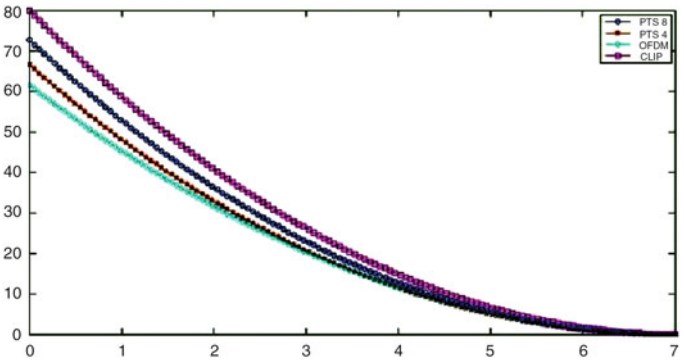


Fig. 5 EE performance with different PAPR reduction at different IBO in OFDM system

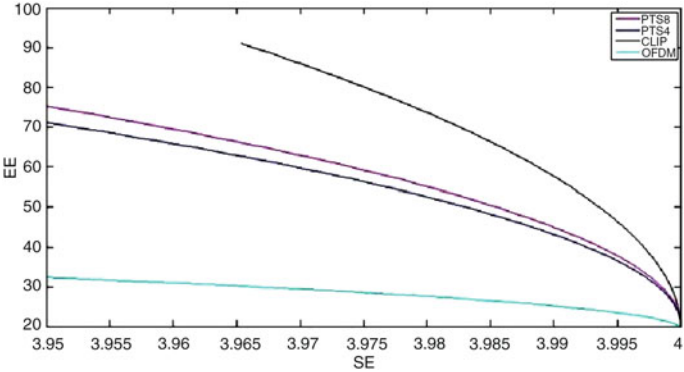
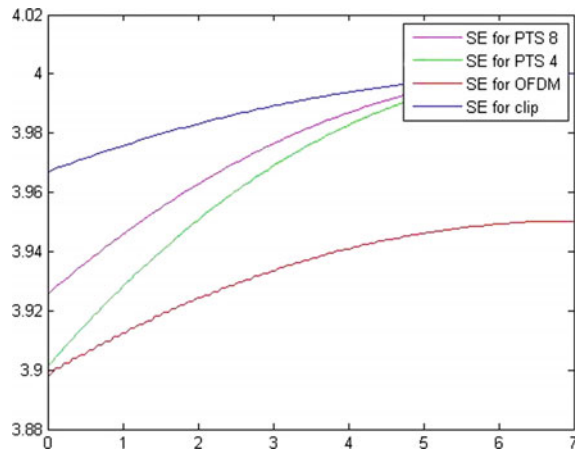


Fig. 6 Relation between SE and EE with constant P_{avg} when PAPR reduction is different

Fig. 7 SE performance with different PAPR reduction at different IBO in OFDM system



6 Conclusion

Thus in the above paper we have analyzed and studied the overall consequence of the PAPR reduction in the SE and EE in OFDM systems considering the CLASS A HIGH POWER AMPLIFIER. With the PAPR reduction, the power efficiency of the HPA is extremely enhanced, and the nonlinear distortion noise caused by the HPA is reduced to remarkable degree. Thus, the results can be obtained with the comparison of the original OFDM scheme without PAPR reduction, the orthogonal frequency division multiplexing systems with PAPR reduction can achieve advanced data rate with very low power consumption. Therefore, both the SE and EE performances can be greatly improved by reducing the PAPR of the OFDM signals. Also PAPR reduction satisfies the requirement for low power in smart devices.

References

1. R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. 4, no. 2, pp. 69–72, Jun. 1958.
2. L. Yang, K. Soo, S. Li, and Y. Siu, "PAPR reduction using low complexity PTS to construct OFDM signals without side information".
3. T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Trans. Broadcast.*, vol. 54, no. 2, pp. 257–268, Jun. 2008.
4. H. Li, T. Jiang, and Y. Zhou, "An improved tone reservation scheme with fast convergence for PAPR reduction in OFDM systems," *IEEE Trans. Broadcast.*, vol. 57, no. 4, pp. 902–906, Dec. 2011.
5. T. Jiang, M. Guizani, H.-H. Chen, W. Xiang, and Y. Wu, "Derivation of PAPR distribution for OFDM wireless systems based on extreme value theory," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 1298–1305, Apr. 2008.

6. T. Jiang, W. Xiang, H.-H. Chen, and Q. Ni, "Multicast broadcast services support in OFDMA-based WiMAX systems," *IEEE Commun. Mag.*, vol. 45, no. 8, pp. 78–86, Aug. 2007.
7. V. Chakravarthy, X. Li, Z. Wu, M. A. Temple, F. Garber, R. Kannan, and A. Vasilakos, "Novel overlay/underlay cognitive radio waveforms using SD-SMSE framework to enhance spectrum efficiency: Part I. Theoretical framework and analysis in AWGN channel," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3794–3804, Dec. 2009.
8. C. Lacatus, D. Akopian, P. Yaddanapudi, and M. Shadaram, "Flexible spectrum and power allocation for OFDM unlicensed wireless systems," *IEEE Syst. J.*, vol. 3, no. 2, pp. 254–264, Jun. 2009.
9. G. Koutitas, "Green network planning of single frequency networks," *IEEE Trans. Broadcast.*, vol. 56, no. 4, pp. 541–550, Dec. 2010. [12] H. Li, T. Jiang, and Y. Zhou, "A novel subblock linear combination scheme for peak-to-average power ratio reduction in OFDM systems," *IEEE Trans. Broadcast.*, vol. 58, no. 3, pp. 360–369, Sep. 2012.
10. J. Chuang and N. Sollenberger, "Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment," *IEEE Commun. Mag.*, vol. 38, no. 7, pp. 78–87, Jul. 2000.
11. Z. Yang, L. Dai, J. Wang, J. Wang, and Z. Wang, "Transmit diversity for TDS-OFDM broadcasting system over doubly selective fading channels," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 135–142.
12. S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.
13. H. B. Jeon, J. S. No, and D. J. Shin, "A new PAPR reduction scheme using efficient peak cancellation for OFDM systems," *IEEE Trans. Broadcast.*, vol. 58, no. 4, pp. 619–628, Dec. 2012.
14. C. Li, T. Jiang, Y. Zhou, and H. Li, "A novel constellation reshaping method for PAPR reduction of OFDM signals," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2710–2719, Jun. 2011.
15. Y. Zhou and T. Jiang, "A novel multi-points square mapping combined with PTS to reduce PAPR of OFDM signals without side information," *IEEE Trans. Broadcast.*, vol. 55, no. 4, pp. 831–835, Dec. 2009.
16. T. Kuroda and M. Hamada, "Low-power CMOS digital design with dual embedded adaptive power supplies," *IEEE J. Solid-State Circuits*, vol. 35, no. 4, pp. 652–655, Apr. 2000.
17. J. Hong, Y. Xin, and P. A. Wilford, "Digital predistortion using stochastic conjugate gradient method," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 114–124, Mar. 2012.
18. S. Shepherd, J. Orriss, and S. Barton, "Asymptotic limits in peakenvelope power reduction by redundant coding in orthogonal frequency division multiplex modulation," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 5–10, Jan. 1998.
19. H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 282–289, Feb. 2001.

Energy Efficient Resource Provisioning Through Power Stability Algorithm in Cloud Computing

Karanbir Singh and Sakshi Kaushal

Abstract Over the past few years energy consumption has become a major operational cost in data centers. Virtualization has been quite instrumental in reducing the energy consumption. Various researches have been focusing on developing energy efficient algorithms for developing power aware resource allocation and scheduling policies. Every virtual machine migration (VMM) incurs extra cost in terms of energy consumption. However, very few techniques exist which particularly focuses on reducing the total virtual machine migrations in a data center. This paper proposes an algorithm which profiles the overall energy consumed based on: max utilization of host after allocation, creation history of virtual machine (VM), and the difference in power consumed by host before and after allocation. The framework for the implementation of the proposed algorithm is conducted in CloudSim. The results show that reducing the total number of virtual machine migrations affects the overall energy consumption in the cloud.

Keywords Virtual machine • Resource provisioning, virtual machine migration • Energy consumption • Stability factor • MBFD

1 Introduction

Cloud computing is one of the biggest changes witnessed by the IT industry in recent times. Cloud computing introduces pay-as-you-go and access-anywhere model. However, modern day data centers continue to grow in complexity and scale. These data centers have become a major consumer of power and energy resources. This consumption results in high operating cost and high carbon

Karanbir Singh (✉) • Sakshi Kaushal
Computer Science and Engineering, University Institute of Engineering and Technology,
Panjab University, Chandigarh, India
e-mail: singh981@gmail.com

Sakshi Kaushal
e-mail: sakshi@pu.ac.in

dioxide emission in the environment. Carbon dioxide emission from datacenters significantly contributes to the green house effect. It contributes around 2 % of the global emission of carbon dioxide [1]. Statistics shows that average energy consumed by each data center is equivalent to energy consumption of 250,000 household appliances. According to survey of American society of heating, refrigerating, and air-conditioning engineers (ASHRAE) in 2014, the infrastructure and energy consumption cost 75 % of the total expenditure, whereas operating a data center costs only 25 % [2]. Power consumption of server is studied in [3] and results show that it cost 7.2 billion dollars in 2005 for the amount of electricity used by servers all over the world. This also includes electricity consumption for cooling purpose and of auxiliary equipments. Facts also indicate that the electricity consumption of this year is double of what it was in 2000. Managing resources in an energy efficient way is the biggest challenge that data centers are facing and it will grow rapidly and continuously unless energy efficient and advance methods of data center management are developed and applied [4–6].

Energy is mainly wasted because computing resources are used inefficiently. According to the previous year's data, even when the servers are rarely at idle mode, the utilization is never 100 % [7]. Servers normally use 10–15 % of their peak capacity but data center owner has to pay expenses of over provisioning which further results in extra Total Cost of Acquisition (TCA) and Total Cost of Ownership (TCO) [7]. Therefore, underutilized servers play a vital role in inefficient energy consumption. Another problem arising due to high energy consumption and increasing number of server components is the heat dissipation. There are efficient cooling systems in today's world but few years back, for 1 watt of power consumed, an additional 0.5–1 W was required for cooling system [8]. Beside the overwhelming cost and electricity bills, another problem arising from this issue is 2 % of global carbon dioxide that is emitted by data centers [1]. According to the estimation by the U.S. Environmental Protection Agency (EPA), the current efficiency trends led to the increase of annual CO₂ emissions from 42.8 million metric tons (MMTCO₂) in 2007 to 67.9 MMTCO₂ in 2011.

All these reasons arise the need of saving energy and power in all aspects and it becomes a first-order objective while designing modern computing systems. The rest of the paper is organized as follows: in Sect. 2, work related to different energy efficient algorithms is discussed. Section 3 describes the problem formulation and the proposed power stability algorithm (PSA). In Sect. 4, we analyzed the proposed algorithm using different parameters. Section 5 concludes the paper.

2 Related Work

Buyya et al. [9] proposed policies for selecting VMs in VM migration. These policies minimize the migration overhead as least number of VMs has been migrated. Cao et al. [10] described an extension of virtual machine consolidation (VMC) policy. In this improved policy, basic MC policy is used. Initially mean and

standard deviation of CPU utilization of host are determined and then further is used to find out whether a host is overloaded or not. Second, on the basis of knowledge of statistics, range of correlation coefficient is divided into negative correlation and positive correlation. Panchal et al. [11] described virtual machine allocation as an important feature in cloud environment and provided information of allocated virtual machine in the datacenter. According to the authors, allocation policies are implemented at infrastructure layer and virtual machine allocation can be made transparent to the user. Wood et al. [12] proposed a hotspot detection algorithm that detects when the VM should be migrated. Greedy algorithm used by hotspot migration determines the destination host for migration as well as evaluates the quantity of resources that need to be allotted to VM after migration. Hai et al. [13] described compression techniques and characteristics-based compression algorithm (CBC/MECOM) for fast, stable live migration of virtual machine data. On source side, data that are to be migrated are compressed first and then migration is done. Ma et al. [14] proposed an improved pre-copy approach. Bitmap page is added to Pre-copy approach, which records or marks the frequently updated pages. Those pages are then added into the page bitmap. So, the updated pages are transferred only once at the end of iterations. This approach minimizes the quantity of data for transferring which further minimizes the total migration time. Using bitmap page also reduces number of iterations. Lie et al. [15] proposed a new approach for virtual machine migration which is known as an improved time series-based pre-copy approach. In this technique, concept of prediction is used to find out those dirty pages that are updated very frequently in the past and a precise prediction is done on those pages that are going to be updated frequently in the future also. Hines et al. [16] proposed post-copy approach for live VM migration. In this approach processor state of VM is first transferred to destination host, started the VM on the destination host and at last the memory pages are transferred. Memory pages that are not successfully transferred are known as demand pages, which are transferred at last from source and then the VM at source is suspended. The main benefit of this approach is that no duplicate transmission of memory pages is done, thus avoiding the overhead for the same as in pre-copy approach. Downtime of post-copy is higher as compared to pre-copy approach.

Stoess et al. [17] proposed a framework for energy management on virtualized servers. Generally, energy-aware OSes have the full knowledge and control over the underlying hardware and based on this, device or application-based accounting is applied in order to save energy. Cardoso et al. [18] deals with the problem of allocating virtual machines in a power efficient way in a virtualized environment. A mathematical formulation of the optimization problem is proposed by the author. Author calculated the power consumption and utility gained from the execution of a VM and named the combined results as “priori”.

From the review of literature, it has been found that there can be further improvement in the allocation policies of VMs to their destination hosts. So in this paper, we proposed a new algorithm which considers the power and stability factor while choosing a destination host. The proposed power stability algorithm (PSA) provides stability while minimizing the overall power.

3 Proposed Work

In cloud computing most of the servers in the data center are running continuously and consuming 70 % of their resources in the idle state [19]. Therefore, it is very difficult to estimate the threshold limits with accuracy as the whole utilization history of the host has to be calculated. If a host has been underutilized for a significant amount of time then it is better to shut it down so as to save energy. But shutting down a host is not easy as just turning a switch on or off. It can lead to SLA violations and maybe a single point of failure and also degrade performance. MBFD algorithm is one of the fastest algorithms available for choosing a destination host [9]. It has a linear complexity. It is used for deciding destination hosts for the purpose of allocation of VMs. It maintains a list a VM that needs to be migrated and a list of destination hosts. MBFD basically maps a VM to its destination host. The work carried out in this paper focuses on the enhancement of allocation policies for VMs such that number of VM migrations and overall energy can be reduced. The proposed algorithm, namely, power stability algorithm (PSA) can successfully reduce the number of migrations and consumption of energy.

3.1 Power Stability Algorithm (PSA)

VM placement and scheduling are studied in the aspects of resource scheduling and VM migration latency. VM allocation in cloud computing environment should be done such that the stability of the destination host is increased, i.e., the host should not be involved in any kind of migration for longer periods of time. To achieve this, we have proposed an algorithm with linear complexity, namely, PSA, which is based on MBFD algorithm. However, the PSA considers a number of factors that has not been considered in MBFD algorithm for selecting best suitable host for a particular VM from the list of all available hosts. The algorithm is based upon the following additional factors:

- Maximum Utilization of host after allocation
- Creation history of the VM
- Power of host after allocation

In general, the process of VM migration consists of the following steps: deciding the instant when to migrate a VM, choosing the most appropriate VM for migration, choosing a destination host where the particular VM shall be migrated, and finally choosing which hosts from the host list need to be switched on/off. Choosing a destination of particular VM is very important. The proposed technique is based on the fact that there is a considerable amount of energy and resources consumed while migrating a VM to a host. Moreover, during migration a particular user may witness degradation in performance. Therefore, we need to minimize the number of migrations so as to improve performance and save energy. This can only be

possible if the stability of the host is increased. Stability of a host means that the total number of migrations in and out of the host is minimum. The more the stability the less the number of migrations. So, less will be the energy consumed in migration of VMs, resulting in overall reduction in energy consumption. We have calculated each host's utilization in our implementation. After calculation of each host's utilization, we made a list consisting of all hosts with their respective utilization values. From this list, while choosing the destination host, we calculated the increase in utilization for each host for that particular VM and selected the host with least increase in utilization.

3.2 *PseudoCode*

This section presents the detailed steps of the algorithm.

Pseudo Code

Begin

Step 1: Get list of all eligible hosts.

Step 2: While host list is not empty, for each host repeat the following:

 Step 2.1: Calculate the maximum utilization of each host for that particular VM.

 Step 2.2: Select host for which increase in utilization is minimum

End loop

Step 3: Check creation history of the VM.

 Step 3.1: If VM is NOT recently created and maximum utilization exceeds upper utilization threshold go to step 2.

 Step 3.2: Else choose that particular host for migration.

Step 4: Obtain the power of host for that particular VM.

Step 5: Calculate power difference of power after allocation and current power of host.

Step 6: Allocate the VM to that host where there is minimum power change.

END

4 Results and Discussions

The proposed algorithm is implemented in CloudSim. In order to analyze the work, various input sets (in terms of tasks/cloudlets) are given to the system. Each cloudlet is created randomly at runtime. Each cloudlet is then added to a central cloudlet list. Similarly, a list of VMs is also prepared at runtime consisting of all the randomly created VMs. Both the cloudlet and VM list are provided to the data center broker at runtime. Datacenters are also created containing the hosts in Cloudsim. In-built functions are used to calculate power of a host for a particular VM and power difference after allocations, etc. The results are analyzed with two

Table 1 Simulation parameters

Number of host machines	200
VM migration lower threshold	20 %
VM migration upper threshold	70 %
Host RAM	16,384 MB
Host bandwidth	10 Gb/s
VM size	2500 MB

performance evaluation parameters, i.e., the total energy consumed and the total number of migrations. The simulation parameters used are shown in Table 1. Five samples are captured by running simulation experiments for particular number of cloudlets.

We have compared our proposed algorithm, i.e., PSA with MBFD algorithm by considering various parameters like energy consumption and number of migrations. The following graphs show the result of the simulations. It is evident from Fig. 1 that energy consumption is increasing with the increase in number of cloudlets. Energy consumption in case of MBFD algorithm is increasing linearly. It can also be noticed that as the number of cloudlets is increasing the difference between both algorithms also increases. This happens because more VMs have to be created to manage the tasks. With increase in number of VMs, the total consumed energy also increases. Between 90 and 100 there is a sharp increase in energy consumption for MBFD algorithm because of random specifications of cloudlets that are being created at periodic intervals. Overall, our proposed algorithm consumed 23 % less energy as compared to MBFD algorithm. The main reason for this difference is that PSA has considered stability factor of a host before migrating a VM to it.

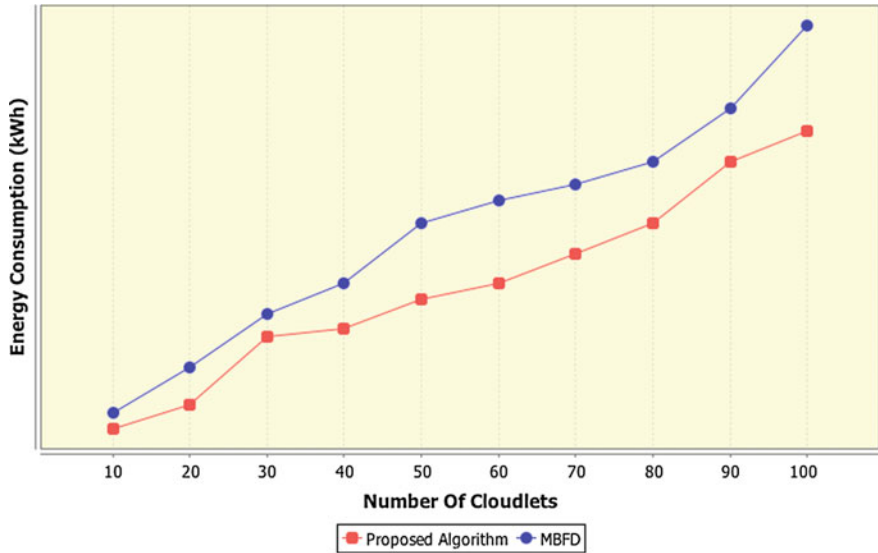


Fig. 1 Energy consumption

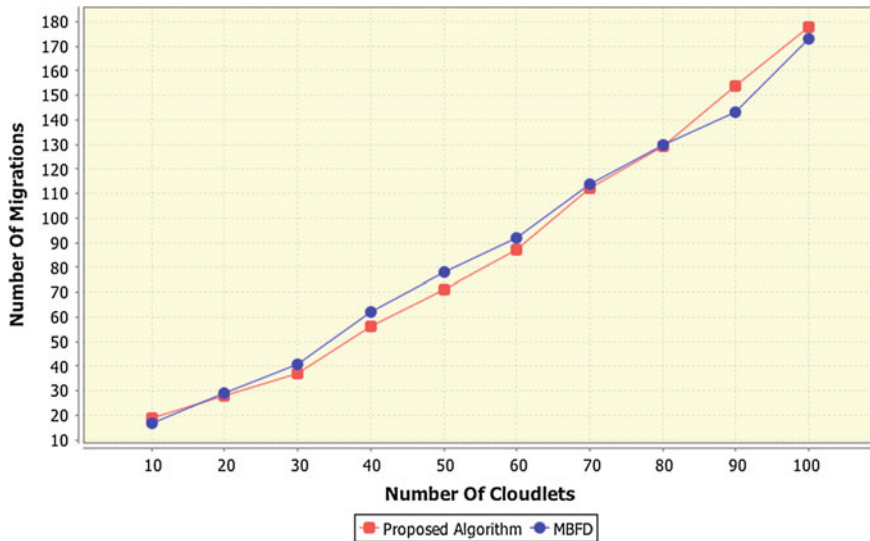


Fig. 2 Number of migrations

As it is observable from Fig. 2 the number of migrations in both the algorithms has been erratic. There is only a minor difference in the number of migrations between both algorithms for a particular number of cloudlets. With increase in number of VMs the total consumed energy also increases. This happens because the proposed algorithm takes into account the stability factor of the destination host. In this algorithm, the VMs have been migrating to hosts having the highest stability factor. This leads to the overall increase in stability of the datacenter, which decreases the total number of migrations that are taking place.

Hence, it is shown that the stability factor of a host has a direct influence on the power consumption of that host. The higher the stability factor less will be the number of migration and therefore less will be the power consumed.

5 Conclusion

This paper focuses on enhancement of VM allocation policies in such a way that energy consumption and number of VM migrations can be reduced. The performance of PSA is evaluated in CloudSim 2.0 simulator for validating the effectiveness and accuracy of results. During each simulation, maximum utilization of host is calculated after each allocation. After each migration the stability factor of host has been recalculated based upon the type of VMs it holds and increase in power after allocation. This strategy proved very efficient in reducing the number of migrations in the data center. As a result PSA consumes 23 % less energy in

comparison with MBFD algorithm. To conclude, the results demonstrate that PSA has immense potential as it offers significant energy saving with comparatively less VM migrations under dynamic workload scenarios as compared to MBFD algorithm.

References

1. C. Pettey, Industry Accounts for 2 Percent of Global CO2 Emissions, www.gartner.com/it/page.jsp?id=503867 (2007).
2. C.L. Belady, "In the Data Center, Power and Cooling Costs More Than the IT Equipment it Supports", *Electronics Cooling*, vol. 13, no. 1, pp. 24 (2007).
3. J. G. Koomey, Estimating total power consumption by servers in the US and the world, Lawrence Berkeley National Laboratory, Tech. Report (2007).
4. "Cloud Computing: Clash of the clouds". *The Economist* (2009).
5. Chaima Ghribi. Energy efficient resource allocation in cloud computing environments. Networking and Internet Architecture [cs.NI]. Institut National des Telecommunications, (2014).
6. Zhihui Lu, Soichi Takashige, Yumiko Sugita, Tomohiro Morimura, and Yutaka Kudo, "An Analysis and Comparison of Cloud Data Center Energy-Efficient Resource Management Technology", *International Journal of Services Computing*, vol. 2, no. 4, pp. 32–51 (2014).
7. Voorsluys, William; Broberg, James; Buyya, Rajkumar, "Introduction to Cloud Computing". New York, USA: Wiley Press. pp. 131–134 (2010).
8. R. Buyya, J. Broberg, A. Goscinski, *Cloud Computing: Principles and Paradigms*. New York, USA: Wiley Press. pp. 1–44 (2011).
9. A. Beloglazov, J. Abawajy and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," *Future Generation Computer Systems*, vol. 28, no. 5, pp. 755–768 (2012).
10. A. Beloglazov and R. Buyya, "Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers," *Concurrency and Computation: Practice and Experience (CCPE)*, Wiley Press, pp. 1397–1420 (2012).
11. M. Nelson, B. H. Lim, and G. Hutchins, "Fast Transparent Migration for Virtual Machines," in *Proceedings of USENIX Annual Technical Conference (USENIX'05)*, pp. 391–394 (2005).
12. R. Buyya, A. Beloglazov and J. Abawajy, "Energy-efficient management of datacenter resources for cloud computing: a vision, architectural elements, and open challenges," in *Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications*, CSREA Press, United States of America, pp. 6–17 (2010).
13. A. Kochut and K. Beaty, "On strategies for dynamic resource management in virtualized server environments," *15th IEEE International Symposium in Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, USA, pp. 193–200 (2008).
14. P. Svard, B. Hudzia, J. Tordsson, and E. Elmroth, "Evaluation of delta compression techniques for efficient live migration of large virtual machines," in *Proceedings of the 7th ACM SIGPLAN/SIGOPS international conference on Virtual execution environments*, ser. VEE'11, ACM, USA, pp. 111–120 (2011).
15. F. Ma, F. Liu, and Z. Liu, "Live virtual machine migration based on improved pre-copy approach," in *IEEE International Conference on Software Engineering & Service Sciences ICSESS*, Beijing, pp. 230–233 (2010).
16. E. P. Zaw and N. L. Thein, "Improved Live VM Migration using LRU and Splay Tree Algorithm," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 3, pp. 1–7 (2012).

17. H. Liu, H. Jin, X. Liao, L. Hu, and C. Yu, "Live migration of virtual machine based on full system trace and replay," in Proceedings of the 18th ACM international symposium on High performance distributed computing, Germany, pp. 101–110 (2009).
18. D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang, "Power and performance management of virtualized computing environments via lookahead control," *Cluster Computing*, vol. 12, no. 1, pp. 1–15 (2009).
19. D. Kliazovich, D. et al., "Accounting for load variation in energy-efficient data centers", *IEEE Conference on Communication*, Budapest, pp. 2561–2566 (2013).

Comparative Analysis of Scalability and Energy Efficiency of Ordered Walk Learning Routing Protocol

Balram Swami and Ravindar Singh

Abstract Mobile ad hoc is the most attractive area of research because of its dynamic topology and mobile environment. It is a structureless self-adjustable network. Routing is the main challenge in MANET. AODV is reactive routing protocol which is based on breadth-first search. DSDV is a proactive routing protocol which maintains a routing table which contains routing information. In this paper, we compare these well-known MANET routing protocol with ordered walk learning routing protocol. OWL is also a reactive routing protocol but it uses DFS in place of BFS. OWL has less congestion than AODV. In this work we propose DOWL and TOWL as two enhancements of basic OWL which uses double DFS and triple DFS instead of single DFS. Both DOWL and TOWL try to minimize the delay and maximize the delivery ratio which consume less energy than AODV, OWL, and DSDV.

Keywords MANET · AODV · DSDV · OWL · DOWL · TOWL · DFS · BFS

1 Introduction

MANET is a structureless network with dynamic topology and it supports mobility of nodes within the network. Because of its dynamic nature and changing topology, routing of packets faces difficulties to perform well. Routing in MANET has two types of routing, first is reactive routing and second is proactive routing. In reactive routing paths are not stored and path discovery is on demand and in proactive routing paths are stored in routing tables which is maintained by each node of the

Balram Swami (✉) · Ravindar Singh
Computer Science Department, Government Engineering College, Ajmer, India
e-mail: mbswami8@gmail.com

Ravindar Singh
e-mail: ravikaviya@gmail.com

network. Nodes of the network can be laptops, cell phones, etc., every node have limited processing power, battery, and bandwidth. Because of MANET's structure and its dynamic nature of topology routing is very challenging. Our main aim of routing is to maximize delivery ratio and minimize end-to-end delay with less energy consumption in routing phases.

AODV is a reactive routing protocol which is based on BFS to discover the path from source to destination. DSDV is a proactive routing protocol which stores the paths in the routing table of the node. This routing table is maintained by every node of the network [1]. OWL is a reactive routing protocol in which each node maintains a routing table which stored the list of neighbor nodes. OWL uses a DFS instead of flooding RREQ message. It uses three kinds of messages to establish communication between the nodes of the network. DOWL and TOWL are the enhancements of the basic OWL. DOWL uses double DFS simultaneously to reduce the delay in route discovery to destination node and TOWL uses three DFS instead of single DFS.

In this paper, we are going to compare and analyze the scalability and energy efficiency of OWL with well-known MANET routing protocols and its enhancements. Section 2 contains the overview of AODV, DSDV, and OWL with its enhancements. Section 3 contains the experimental results of AODV, DSDV, and OWL. Section 4 contains the conclusion of the paper and future scope of the paper.

2 Routing Protocols Overview

2.1 AODV

It is a reactive routing protocol [2]. It uses BFS to find route from source node to the destination node on demand. AODV uses RREQ, RREP, RERR control messages to establish communication between the nodes of the network. It can be used for large network in which there are more than thousands of nodes. To discover route it will flood the network by RREQ message. The main disadvantage of flooding is congestion on the network because of RREQ control messages, if several nodes broadcast RREQ at the same time [3]. This will decrease the delivery ratio and increase end-to-end delay of protocol. If there is large end-to-end delay of packets than threshold time delay then packets will be dropped within the network and due to this delivery ratio decreases because large number of packets are dropped from total generated packets. AODV provide loop-free route while repairing broken links [2] (Fig. 1).

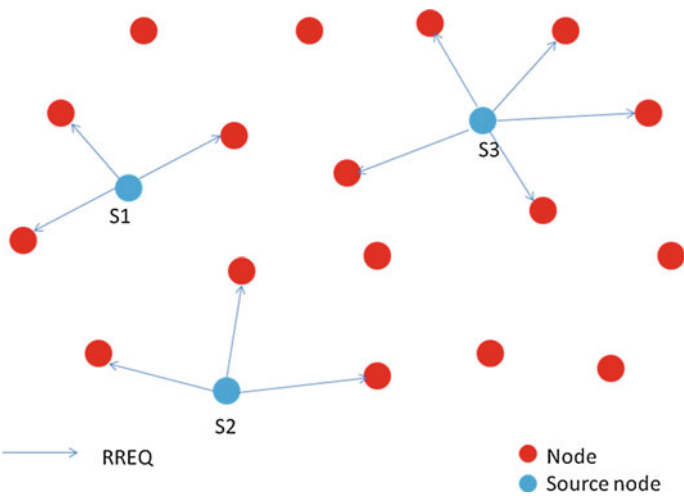


Fig. 1 Shows the working of AODV, source nodes broadcast RREQ to all of its neighbors

2.2 DSDV

Destination-sequenced distance vector routing protocol was developed by Perkins and Bhagwat in 1994 based on Bellman–Ford algorithm [1]. In DSDV each node maintains a routing table which contains the routing information about the network like destination id, hop count, unique sequence number, etc. Every node of the network knows about the structure of the network [1]. Every node forward its own routing table to all of its neighbor nodes after a particular time interval or it may be based events (when a new node is joining or delete existing node). Table forwarding is done by broadcasting or multicasting to all the neighbors of the node (Fig. 2). The entire node either forwards complete table “full dump [4]” or just forwards the updates made in its routing table “increments [4]” to the neighbors (Table 1).

Table 1 Shows the routing table of node A in the network [4]

Destination	Next hop	Number of hops	Sequence number	Install time
A	A	0	A46	001000
B	B	1	B36	001200
C	B	2	C28	001500

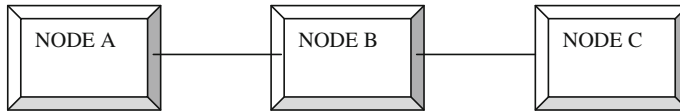


Fig. 2 Shows the frequent network in which DSDV is used for routing

2.3 OWL

OWL is a reactive routing protocol which discovers route from source to destination on demand and it uses three types of messages to communicate between nodes of the network (Fig. 3).

- RREQ—route request message generated by the source node.
- RERR—route error message generated by leaf node on failure of route searching and the RERR send back to the source node to acknowledge the source node about route failure and starting the another DFS.
- RREP—route reply message generated by the destination node when a RREQ message is arrived.

OWL uses DFS instead of flooding [5] to minimize the delay and maximize the use of node's resources. In AODV every node floods the network with RREQ messages when discover route from source node to the destination node. This will leads to increase in congestion on the network and it will increase the delay. Because of longer delay some packets are dropped in between the links due to time out and it will decrease the delivery ratio. OWL increases the use of bandwidth and delivery ratio using DFS. OWL is also an energy-efficient reactive routing protocol.

2.4 DOWL

DOWL stands for double-ordered walk learning routing protocol, which is an enhancement of OWL [6]. It uses double DFS simultaneously instead of single DFS. DOWL tries to minimize the delay using double DFS (Fig. 3).

2.5 TOWL

TOWL stands for triple-ordered walk learning routing protocol, which is also second enhancement of OWL. It uses triple DFS simultaneously instead of single DFS. TOWL tries to minimize the delay using triple DFS (Fig. 3).

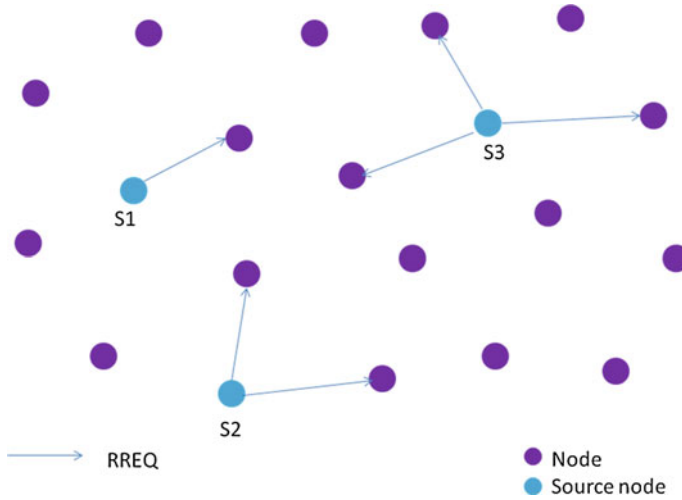


Fig. 3 Shows the working of OWL, DOWL, and TOWL. In the *above figure*, S1 is working on OWL, while S2 and S3 uses DOWL and TOWL, respectively. S1, S2, and S3 are the source nodes in the networks

3 Simulation Results

Experimental results are based on the energy consumption on different phases of the routing.

- Routing energy consumption—total energy consumed in receiving and forwarding of packets at node (Fig. 4)
- Data energy consumption—total energy consumed in routing of packets and transmission of packets (Fig. 5)
- Overall energy consumption—total energy consumption in routing in different steps of routing (Fig. 6).

3.1 Scalability

Scalability of routing protocols is the ability of a routing protocol to work well in large-scale network [7]. AODV has high scalability than OWL and DSDV, because it has higher delivery ratio and low delay than OWL and DSDV. DSDV has very low scalability because of routing tables, maintaining and synchronization of tables causes very high overhead and consume node's resources more than AODV and OWL. OWL has moderate scalability because it works worst than AODV but it can perform comparable if network has low mobility of nodes [8].

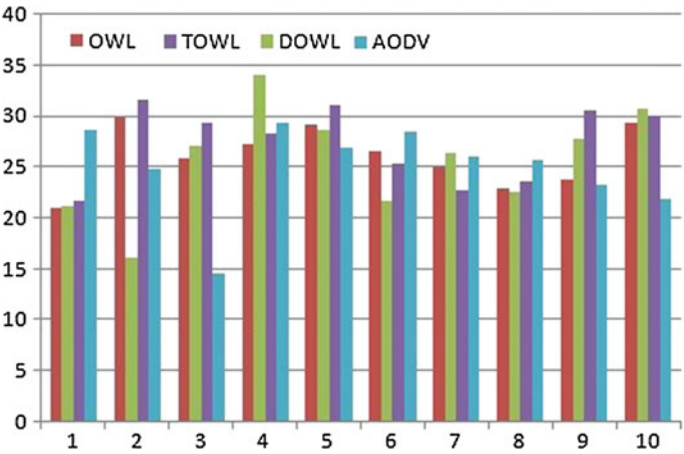


Fig. 4 Shows the routing energy consumption of OWL, DOWL, TOWL, and AODV. In the *above figure* the routing energy of DSDV is not displayed because its routing energy in large networks is so high than AODV and OWL

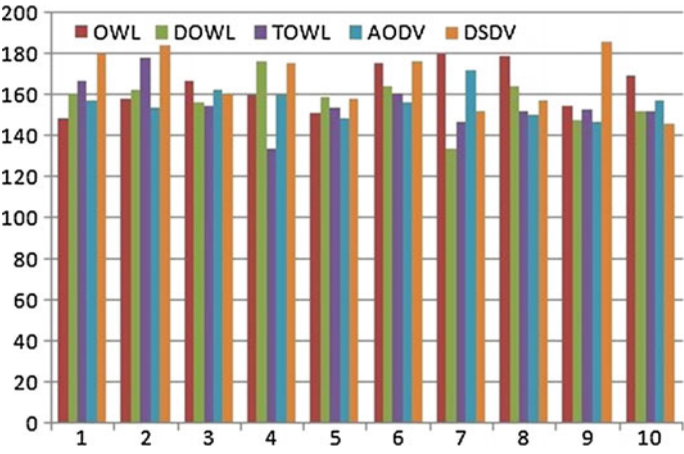


Fig. 5 Shows the routing energy consumption of OWL, DOWL, TOWL, AODV, and DSDV. DSDV stores path in routing tables so DSDV also have comparable energy consumption with AODV and OWL in large network

3.2 Energy Efficiency

Energy efficiency of routing protocol is how a routing protocol is efficient with respect to the consumption of energy of nodes of the network, because each node of the network has a limited energy. OWL is most energy efficient in moderate size of network and in small-scale network DSDV also has good energy efficiency but in large-scale network AODV have highest energy efficiency than OWL and DSDV [9, 10].

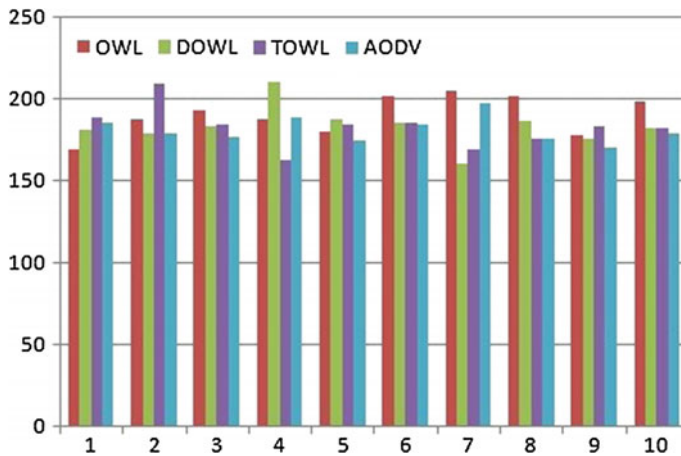


Fig. 6 Shows the total energy consumption in packets transmission and packet routing at nodes of the network by OWL, DOWL, TOWL, and AODV. In the *above figure*, we exclude the results of DSDV because DSDV has very high overall energy consumption than AODV and OWL

4 Conclusions and Future Scope

AODV has very high scalability than OWL and DSDV. OWL works well in moderate size of network and DOWL and TOWL have high scalability than OWL but have low energy efficiency than OWL and AODV in large-scale network. DSDV is not suitable for large-scale network because of high overhead of routing table maintenance. DOWL and TOWL are enhancements of basic OWL which are trying to minimize delay but they consume more energy than OWL. DOWL and TOWL have high scalability than OWL. But still there is a way to maximize the scalability and energy consumption using priority to the nodes. Past information can be used to increase the efficiency of OWL. OWL is not completely explored, still it needs more work to exploit the advantage of OWL.

References

1. Perkins, Charles E. and Bhagwat, Pravin "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers". 1994.
2. Charles E. Perkins & Elizabeth M. Royer "Ad-hoc On-Demand Distance Vector Routing". Proceeding of 2nd IEEE Workshop on mobile computing systems and applications, pages 90–100, New Orleans, LA, Feb 1999.
3. Perkins, C.; Belding-Royer, E.; Das, S. "Ad hoc On-Demand Distance Vector (AODV) Routing". July 2003.
4. Nayan Ranjan paul & Laxminath Tripathy & Pradipta Kumar Mishra "Analysis and Improvement of DSDV". Volume 8, Issue 5 IJCSI September 2011.

5. Stephen Dabideen & J.J. Garcia-Luna-Aceves "OWL: Towards scalable routing in MANETs using depth-first search on demand". Sixth IEEE international conference on mobile ad-hoc and sensor systems 2009.
6. Stephen Dabideen & J.J. Garcia-Luna-Aceves "Efficient routing in MANETs using ordered walks". Springer Science Business Media, 2011.
7. Sivapriya G & Senthil Kumaran T "Smart Route Discovery in MANET using Ordered Walk". International Journal of Computer Applications (0975-8887) Volume 70, No. 4, May 2013.
8. Charles E. Perkins & Elizabeth M. Royer "An Implementation Study of the AODV Routing Protocol" IEEE 2000.
9. S. Shenakaran & A.Parvathavarthim "An Overview of Routing Protocols in Mobile Ad-Hoc Network". IJARCSSE Volume 3, Issue 2, February 1013.
10. Anju Gill & Chander Diwaker "Comparative Analysis of routing in MANET". IJARCSSE Volume 2, Issue 7, July 1012.

A Novel Technique for Voltage Flicker Mitigation Using Dynamic Voltage Restorer

Monika Gupta and Aditya Sindhu

Abstract This paper deals with mitigation of voltage flicker using an intelligent dynamic voltage restorer (DVR). Voltage flicker is produced in the distribution system due to an arc furnace which is a highly nonlinear load in nature. The control scheme of the proposed DVR is based on a neural network (NN) controller whose weights are trained using hybrid of particle swarm optimization (PSO) and gradient descent (GD). A comparative analysis is done for three different controllers: proportional integral (PI), NN with GD, and NN with hybrid of PSO and GD. Simulated results based on peak overshoot and maximum percentage total harmonic distortion (THD) of load voltage shows the superiority of the proposed NN controller.

Keywords Voltage flicker · Arc furnace · Particle swarm optimization · Gradient descent · Neural network

1 Introduction

With the widespread grid integration of renewable energy and increased transmission of power to rural as well as geographically remote areas over the past few years, the prominence of improving the power quality at both the load and generation end has risen substantially. However, the increased use of nonlinear loads and external factors has also led to a rise in deterioration of power quality and introduction of undesirable phenomenon like voltage sag, voltage swell, harmonics, and induced voltage flicker [1]. An unbalanced system consists of displaced amplitudes as well as phase angles in one or all three phases, usually caused by

Monika Gupta (✉) · Aditya Sindhu
Department of Electrical & Electronics Engineering,
Maharaja Agrasen Institute of Technology, Delhi, India
e-mail: gupta.monika.219@gmail.com

Aditya Sindhu
e-mail: aditya.sindhu@hotmail.com

induced faults [2]. Phenomenon like unsymmetrical faults and voltage flickers pose as serious problems for grid safety [3].

Voltage flicker is an uncharacterized sharp change in voltage level accompanied by increased harmonic distortions and is commonly observed in a highly inductive load-like arc furnace. Although voltage flicker lasts for a short-time burst, it is considered potentially hazardous for the load and the supply conductors as well. Voltage flicker is primarily caused by load end reactive power variations, a characteristic commonly shown by highly nonlinear loads-like arc furnace and induction heating [4]. This not only poses a threat to the supply system infrastructure but also jeopardizes the safety of appliances running on the same supply. For this reason, its quick detection and mitigation is of utmost importance.

For mitigation of the above potentially hazardous phenomenon and specifically voltage flickers, FACTS devices such as dynamic voltage restorer (DVR) are extensively used in the power industry. A DVR is a voltage mitigation power electronics device that injects the required corrective voltage whenever it senses that the supply or load end voltage level has increased or decreased beyond a set acceptable threshold level. The basic principle of DVR's working involves a control scheme which compares the input voltage level with the set level and subsequent signal generation to the voltage source inverter (VSI). The robustness and efficiency of the DVR's performance thus rely heavily on its control strategy.

In this paper, we have modeled a DVR connected to an arc furnace which is a highly nonlinear load and measured the performance of the DVR individually in MATLAB SIMULINK environment with three different controllers—neural network (NN) controller, the traditionally used PI controller and a NN controller whose weights have been trained by hybrid of particle swarm optimization (PSO) and gradient descent (GD). PSO being a swarm-based algorithm is a comparatively more reliable and robust algorithm with negligible chances of occurrence of local minima [5]. A comparative analysis of the DVR performance is then done for the three controllers and is consequently discussed.

The paper organization is as follows: The DVR operation and its control scheme are discussed in Sect. 2. Modeling of the arc furnace in the SIMULINK environment is discussed in Sect. 3, followed by simulation results in Sect. 4. In Sect. 5 comparison of the controllers is done followed by conclusion in Sect. 6.

2 Dynamic Voltage Restorer (DVR)

The DVR is a FACTS device which essentially injects corrective voltage to mitigate any voltage unbalance in the system, as shown in Fig. 1. It consists of an inverter whose input is controlled by a PWM generator which is further directed by a control strategy. DVR is usually connected at the load end of the supply utility and is connected to the supply via a multiple arm injection transformer.

Whenever the voltage levels increase or decrease beyond a set voltage, gated signal is send to the input of the PWM generator whose function is to produce a

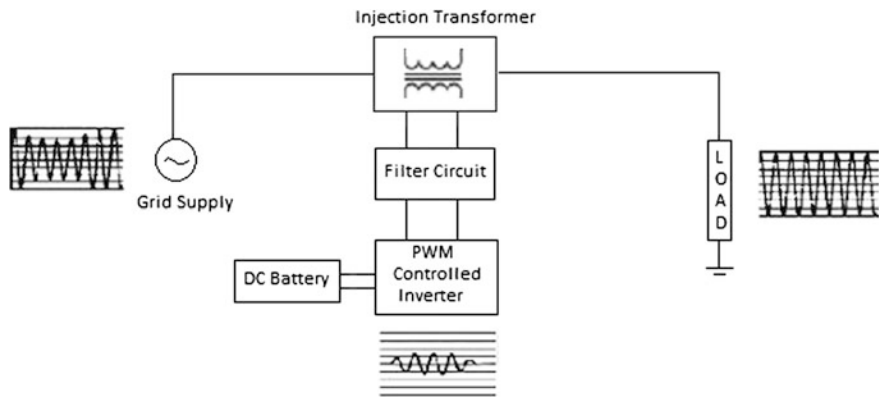
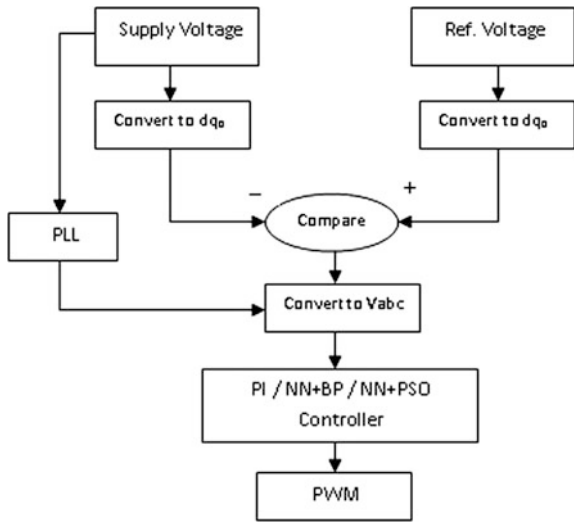


Fig. 1 Basic DVR model

Fig. 2 Control scheme for the DVR using PLL and controller block



rectangular-pulsed waveform responsible for generation of AC voltage by the DC to AC inverter. In mostly observed cases the sag/swell margins vary from 10 to 90 percent, however this may vary greatly in the case of voltage flicker and depending upon load to load and system to system a suitable threshold level V_{ref} can be set.

The controller block in the DVR control strategy is responsible for comparing the set threshold voltage level with the load voltage (V_{load}) and generating the signal for PWM. The flowchart of basic control scheme for the DVR modeled in this paper is given in Fig. 2.

In this paper we have used a multilayered feed forward artificial neural network (MLFF) as a controller for the DVR. The efficiency of the neural network chiefly relies on the weight training algorithm used. In this paper the weight training

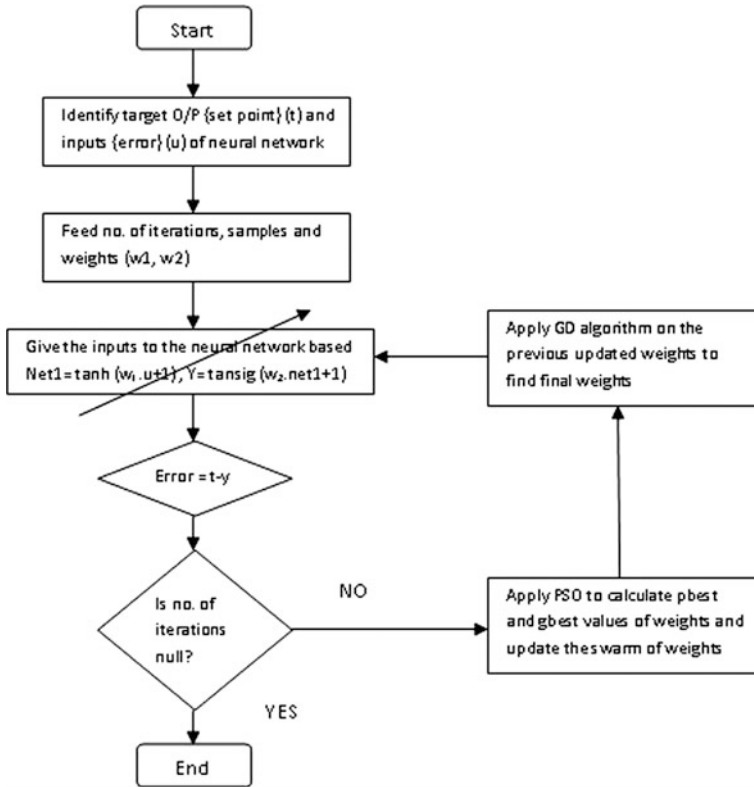


Fig. 3 Flowchart for the hybrid NN controller

algorithm used is a hybrid of PSO and gradient descent (GD). GD is a fast and efficient computational technique but we have not used GD alone as it is not as robust as PSO and there are also chances of getting stuck in local minima, which is not the case with PSO. The drawback of PSO (that it is slow) can be eliminated with using both PSO and GD with the NN controller. First weights of NN are trained using PSO and then GD is applied to evaluate final weights in each iteration as shown in Fig. 3. We have briefly described the PSO algorithm as follows.

PSO involves initialization of a swarm of particles with random velocities and distances in the sample space with the intent of reaching a particular coordinate in space. The result of each iteration is then compared with set parameters Gbest and Pbest. The velocity and distance equations for the swarm of particles are given as follows:

$$v(t + 1) = wv(t) + c_1r_1[\hat{x}(t) - x(t)] + c_2r_2[g(t) - x(t)] \quad (1)$$

$$x(t+1) = x(t) + v(t+1) \quad (2)$$

where x is position of particle at time t , v is the velocity of particle at time t , c_1 and c_2 are acceleration constant for cognitive and social component, respectively, and r is the stochastic constant.

3 Modeling of Arc Furnace

Figure 4 shows the SIMULINK implementation of the DVR model connected to the arc furnace.

The generator rated 110 V, 15 MVA (60 Hz) is connected to an arc furnace (highly nonlinear load) and the DVR is connected via an injection transformer at the point of common coupling. The components of the DVR include the VSC inverter, the PWM generator, and the control subsystem consisting the controller. The electric arc furnace modeling, in practice involves the following six steps:

- After initial charging of electrodes they are brought over the slag. At this point current starts flowing
- Molten metal formation process started
- Arc reaches maximum length, Voltage at it's peak
- Arc is shortened for maximum heat exposure to slag
- Steel refining processes carried out on the molten steel
- Melting process halted.

Evidently, the steps of physical welding process involve a wide fluctuation in load at each step, making this a highly nonlinear load. In order to model this in SIMULINK environment, we implemented a Thevenin equivalent of each stage and coordinated them with successively timed breakers. Figure 5 shows the SIMULINK model for an electric arc furnace.

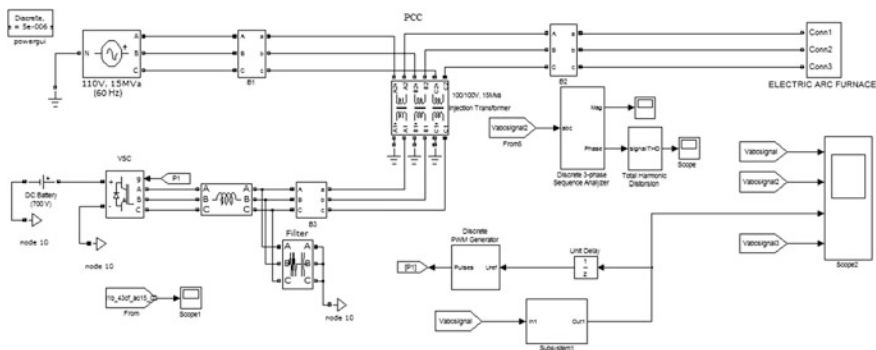


Fig. 4 SIMULINK model for DVR with arc furnace as load

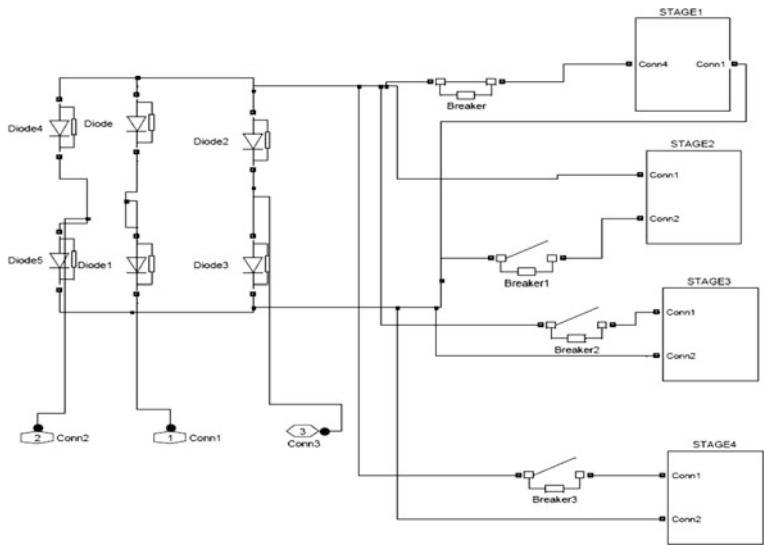


Fig. 5 Arc furnace as implemented in SIMULINK

4 Simulated Results

Figure 6 shows the load (arc furnace) voltage and current waveforms measured at the PCC. Figures 7, 8 and 9 display the waveforms corresponding to the three controllers- PI, NN, and hybrid NN with PSO and GD, respectively. In all three cases, the first graph corresponds to the supply voltage; the second corresponds to the load voltage, and the third to the output of the DVR.

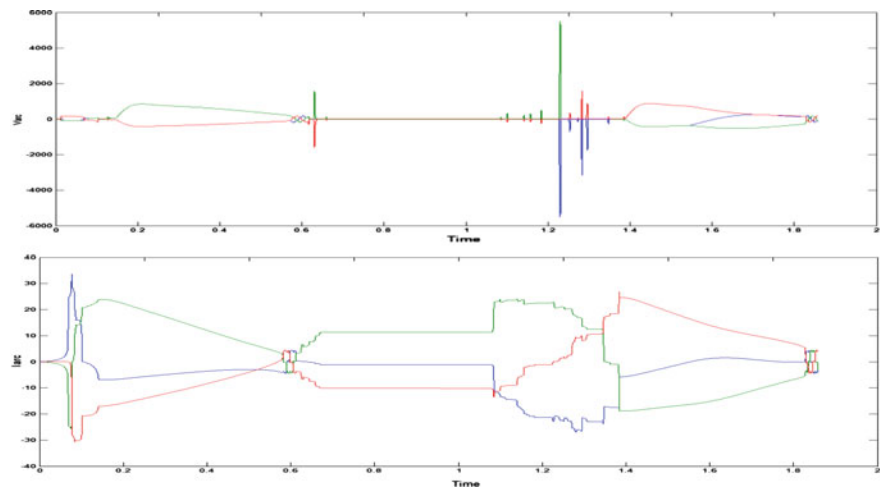


Fig. 6 Plots for V_{load} and I_{load} without DVR

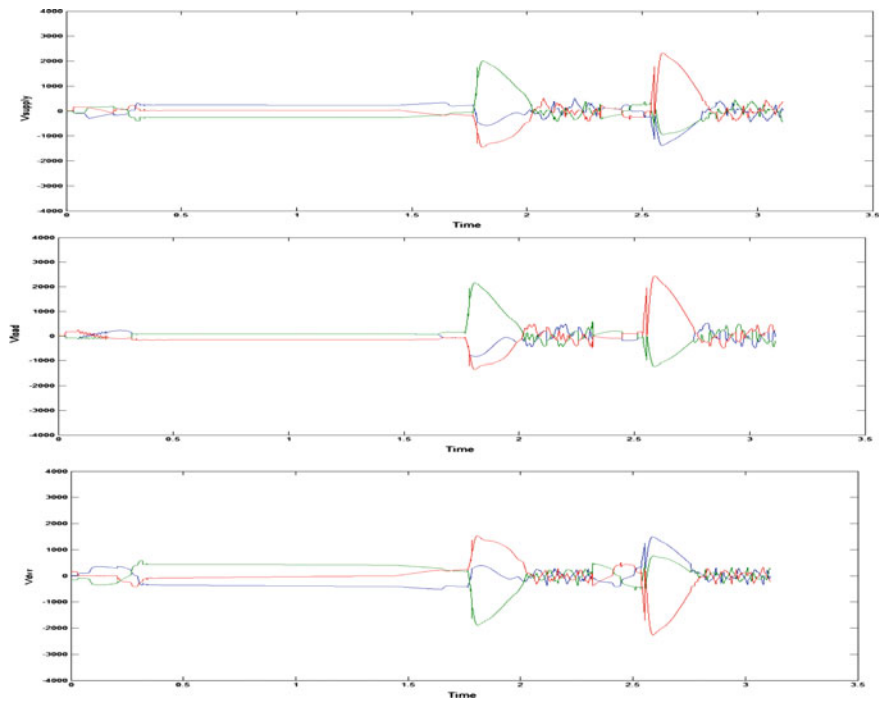


Fig. 7 Plots for V_{supply} , V_{load} and V_{dvr} with DVR (PI controller)

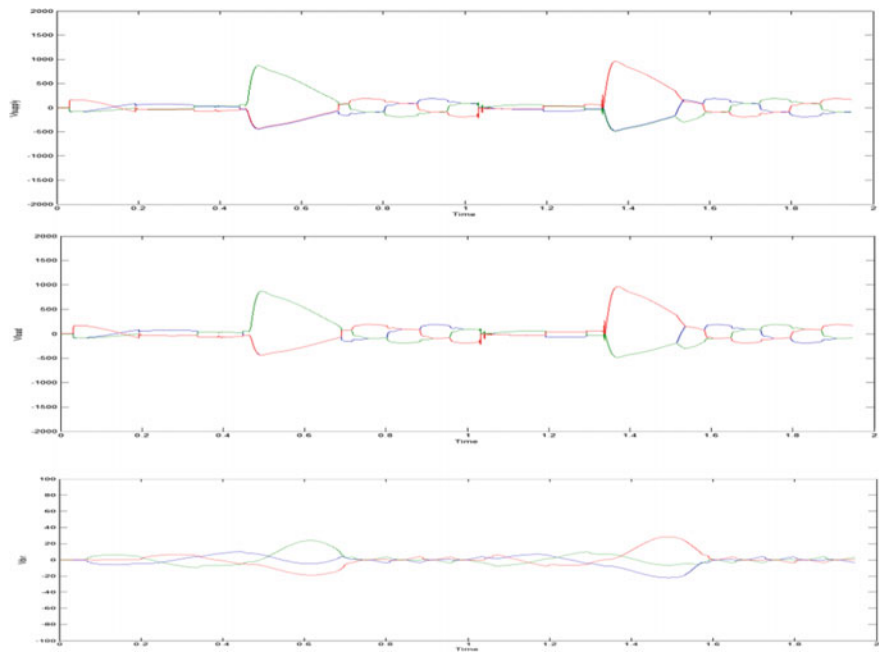


Fig. 8 Plots for V_{supply} , V_{load} and V_{dvr} with DVR (NN controller)

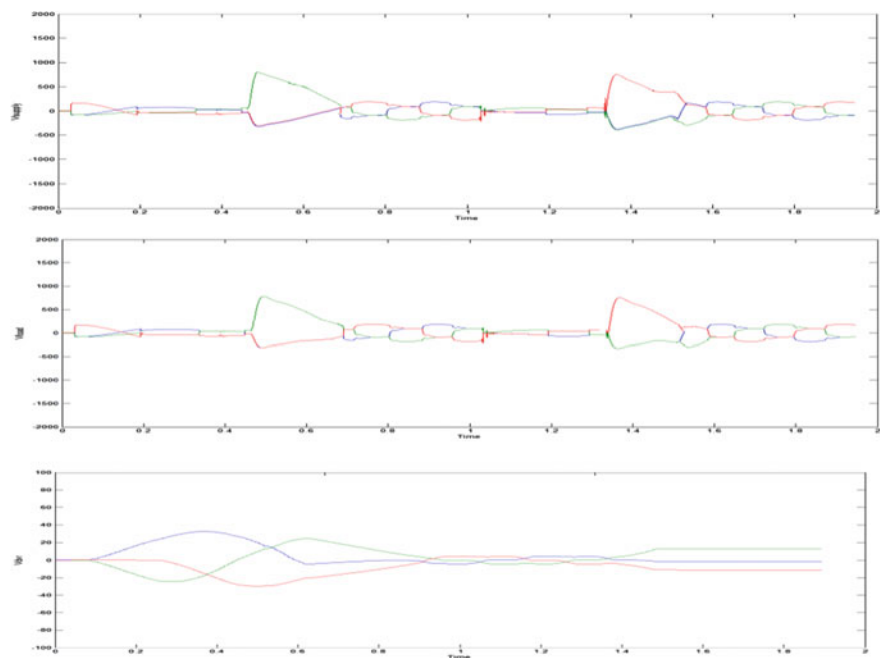


Fig. 9 Plots for V_{supply} , V_{load} and V_{dvr} with DVR (NN+PSO+GD controller)

Table 1 Peak overshoot and Maximum THD values for load voltage with different controllers

Type of controller	Peak overshoot (V)	Maximum THD (%)
PI	2300	5.4
NN+GD	940	4.4
NN+PSO+GD	750	3.7

5 Comparative Analysis of the Controllers

Referring to Figs. 7, 8 and 9 Table 1 has been tabulated. As clear from Table 1, peak overshoot of the hybrid (PSO+GD) controller is the least among the three controllers and it also has the least THD. Maximum THD is the percentage THD measured after the first initial transient, once the voltage level has been steadied [6].

6 Conclusion

In this paper, a comparative analysis of DVR's performance is done for voltage flicker mitigation with different controllers. The voltage flicker is induced by an arc furnace. The controllers under study are PI, NN (GD), and NN (PSO+GD). From

the simulation results it is affirmed that the performance of the proposed NN controller is superior to the others. This work can be extended to its hardware implementation.

References

1. Hocine, L., Yacine, D., Kelaiaia, S., Mounia, K., Bounaya: Closely Parametrical Model for an Electrical Arc Furnace. In: Proceedings of World Academy of Science, Engineering and Technology, Volume 30 (2008).
2. Sankaran, C.: Power quality. CRC PRESS (2002).
3. IEEE recommended practices and requirements for harmonic control in electrical power system. IEEE Standard 519–1992 (1992).
4. Srivastava, S., Gupta, M., Preeti, K.S., Jyoti, D.: Online Estimation and Control of Voltage Flicker Using Neural Network. In: TENCON 2009 - 2009 IEEE Region 10 Conference, pp. 1–6 (2009).
5. Choi, S.S., Li, B.H., Vilathgamuwa, D.M.: Dynamic voltage restoration with minimum energy injection. In: IEEE Trans. Power Syst., vol. 15, no.1, pp. 51–57 (2000).
6. Aali, S., Nazarpour, D.: Voltage Quality Improvement with Neural Network-Based Interline Dynamic Voltage Restorer. In: Journal of Electrical Engineering & Technology Vol. 6, No. 6, pp. 769–775 (2011).

Gaussian Membership Function-Based Speaker Identification Using Score Level Fusion of MFCC and GFCC

Gopal, Smriti Srivastava, Saurabh Bhardwaj
and Preet Kiran

Abstract In this work, a speaker identification system is employed using mel-frequency cepstral coefficients (MFCC) and gammatone frequency cepstral coefficients (GFCC) features. MFCC is the most common feature extraction technique used in speaker identification/verification system and gives high performance in clean environmental conditions. GFCC is known for its noise robustness performance and is highly suitable in noisy or office environment conditions. Here, we combine the advantages of both the feature extraction techniques by score level fusion. Also, we employed a more simpler Gaussian Membership Function (GMF) based matching process. Lastly, we use k-Nearest Neighbor (KNN) to measure the similarity in the verification stage. Experimental results verify the validity of our proposed approaches in personal authentication.

Keywords Speaker identification • MFCC Features • Multimodal system • Score level fusion • Gaussian membership function

1 Introduction

Biometrics can be physiological characteristics or behavioral characteristics. It includes iris recognition, speaker identification, fingerprint identification, hand geometry, face geometry, and several others. MFCC has been the most popular

Gopal (✉) · Smriti Srivastava
Netaji Subhas Institute of Technology, New Delhi, India
e-mail: gopal.chaudhary88@gmail.com

Smriti Srivastava
e-mail: smriti.nsit@gmail.com

Saurabh Bhardwaj · Preet Kiran
Thapar University, Patiala, India
e-mail: bsaurabh2078@gmail.com

Preet Kiran
e-mail: preetarora.1206@gmail.com

feature extraction technique over the last many decades. These are spectral-based features acquired by direct application of the Fourier transform (or fast Fourier transform (FFT) or short-time Fourier transform (STFT)), converted into more robust, flexible, and highly decorrelated and compact representation of cepstral coefficients with the use of perceptual-based mel-spaced filter bank followed by discrete cosine transform (DCT) [1]. MFCC outperforms linear prediction cepstral coefficients (LPCC) in most of the problems, but under clean and matched conditions only and has low robustness to noisy and mismatched conditions. In MFCC, noise cannot be removed from the portions where it overlaps the signal spectrum. Also, this noise corrupts all the frequency bands of speech, because discrete cosine transform cover all frequency bands which affects all the coefficients of MFCC. A frame of speech may contain information of two phonemes while MFCC is inherited to one phoneme at a time in a speech frame.

The human ability to perform speaker recognition in noisy conditions has motivated studies of robust speaker recognition from the perspective of computational auditory scene analysis. GFCC has nonlinear frequency distribution characteristics that have significant advantages in its noise robustness and is free from harmonic distortion and computational noise. Unimodal biometric system is based on a single trait and it suffers from various limitations such as spoof attacks and several others as stated in literature [2–4]. While a multimodal biometric system is created by fusing various unimodal systems to ensure the high performance of such biometric system as the evidences from different sources are combined together to avoid limitations of one. In this paper, we have used the advantages of multimodal biometric system by combining the MFCC and GFCC features by score level fusion.

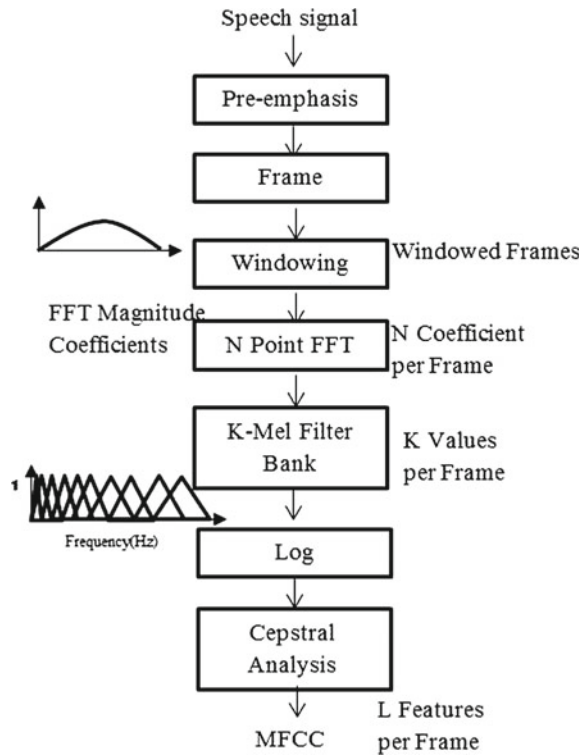
This paper is organized as follows. Section 2 consists of feature extraction in which mel-frequency cepstral coefficients and gammatone frequency cepstral coefficients are discussed. Sections 3 and 4 consists of discussion on score fusion and their rules. Section 5 consists experiments and results and concluded finally in Sect. 6.

2 Feature Extraction

2.1 Mel-Frequency Cepstral Coefficients

Referring to Fig. 1, MFCC features are derived as follows: First, the continuous speech signal is divided into frames of N samples, with adjoining frames separated by M samples. Next, a Hamming window is used to partition each frame, and then a fast Fourier transform (FFT) is applied. The mel-frequency scale corresponds to a linear scale and is given in Eq. 1.

Fig. 1 Analysis block diagram for MFCC feature vectors



$$\text{Mel}(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

The output evaluated after K Mel filter bank $H_m[k]$ is denoted as S_k in Eq. 2.

$$S_k = \sum_{k=0}^{N-1} (|X[k]|^2 H_m[k]) \quad 0 \leq m < K \quad (2)$$

The logarithm of the filter bank output, $\log(S_k)$, is usually taken to reflect the logarithmic compression of human hearing. The final step is to convert the K log filter bank spectral values into L cepstral coefficients using the discrete cosine transform as given in Eq. 3.

$$C_n = \sum_{k=1}^K \log(S_k) \cos \left(n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right) \quad (3)$$

2.2 Gammatone Frequency Cepstral Coefficients

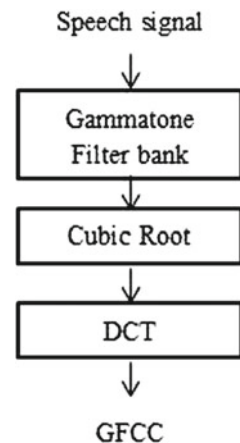
First, auditory filtering is performed on the noisy and reverberant speech by decomposing an input signal into T-F domain via passing the input signal through a gammatone filter bank derived from psychophysical and physiological observations of the auditory periphery to create a two-dimensional cochleagram. A cochleagram gives a finer frequency resolution at low frequencies, based on equivalent rectangular bandwidth (ERB) scale, than at high frequencies than the linear frequency resolution (mel scale) of a spectrogram. In this process, the filter output still retains its original sampling frequency. Thus, this response is decimated to 100 Hz in time domain and results into a frame rate of 10 ms. This frame rate is common in many speech feature extraction methods. Then a cubic root operation is performed on the decimated outputs to generate a gammatone feature (GF) vector as in Eq. 4:

$$G_m[i] = |g|_{\text{decimate}}[i, m]^{1/3} \quad (4)$$

$i = 0, 1, \dots, N - 1, m = 0, 1, \dots, M - 1$. Here, N refers to the number of frequency (filter) channels. M is the obtained decimated time frames. The resulting output is in the form of a matrix which represents the T-F decomposition of the input. In size, GF vector is greater than that of MFCC vectors used in a classic speaker recognition system. Also GF components are highly correlated with each other due to the frequency overlapping. In order to reduce dimensionality and decorrelate the components, we apply DCT to GF. Then, we apply discrete cosine transform to GF to derive GFCC as in Eq. 5.

$$C[j] = \sqrt{\frac{2}{N}} \sum_{i=0}^{N-1} G_m[i] \cos\left((2i+1) \frac{j\pi}{N}\right) \quad (5)$$

Fig. 2 Analysis block diagram for GFCC feature vectors



Detailed feature extraction can be found in [5]. Referring to Fig. 2, GFCC [6] features are derived as follows:

Then the gaussian membership function-based feature is extracted from both MFCC and GFCC features. The feature vector so obtained has a length of 100. The Gaussian membership function (GMF) [7, 8] based feature extraction to extract GMF feature a_i from i th window can be expressed in Eqs. 6 and 7,

$$u_i = \frac{\exp - (x_k - \bar{x})^2}{2\sigma^2} \quad (6)$$

$$a_i = \frac{1}{K} \sum_{i=0}^K x_i u_i \quad (7)$$

where x_k is the feature value at k th point of the window, $\bar{x}$ is mean feature value and σ is the standard deviation of the window, u_i is the Gaussian membership function and a_i is the feature obtained from the i th window.

3 Score Level Fusion of Two Biometrics

The score level fusion also called as confidence level fusion refers to combining the matching scores obtained from different classifiers. The block diagram depicting score level fusion is shown in Fig. 3. Each biometric modality provides a similarity score indicating the proximity of the test feature vector with the template feature vector. The fusion at score level is the most appropriate approach to multimodal biometrics and is most popular. The advantages of score level fusion are as follows: The matching scores (genuine and imposter) from the existing and proprietary unimodal systems can be easily utilized in a multimodal biometric system. The information (i.e., the match score) from prior unimodal evaluations of a biometric system can be used and this avoids live testing. The matching scores contain next level of rich information after the features of the input pattern. The scores generated by different matchers are easy to access and combine. This motivates combining information from individual biometric modalities using score level fusion. *Min-Max score normalization* is done for making combination meaningful [9]. Let r_k denote a set of matching scores, where $k = 1, 2, \dots, n$ and $r_{k'} = \frac{r_k - \min}{\max - \min}$ which denotes normalized score.

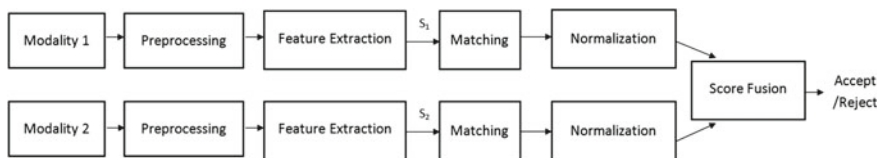


Fig. 3 Score level fusion

4 Score Fusion Rules

Various conventional and t -norm-based fusion rules are given below. Let R_i be the matching score obtained from i th modality and R denotes the fused score or the combined score and N be the number of modalities.

- 1. Sum rule: $R = R_1 + R_2 + \dots + R_N = \sum_{i=1}^N R_i$
- 2. Product Rule: $R = R_1 * R_2 * \dots * R_N = \prod_{i=1}^N R_i$
- 3. Hamacher t -norm: $R = \frac{R_1 R_2 R_3}{R_1 + R_2 + R_3 - R_1 R_2 - R_3 R_2 - R_1 R_3 + R_1 R_2 R_3}$
- 4. Frank t -norm: $R = \log_p \left(\frac{1 + (p^{R_1} - 1)(p^{R_2} - 1)(p^{R_3} - 1)}{p - 1} \right)$

5 Experiments and Results

Twenty two-dimensional MFCC and GFCC, with 0th coefficient removed, are used for this study on VoxForge speech corpus. The data set used for testing is mixed with noise signals at different SNRs (5, 5, 10, and 20 dB). The noise signals are taken from the database NOISEX [10]. To calculate the scores between the training and test sample, the k-Nearest Neighbor (KNN) classifier with Euclidean distance is trained with features obtained from each biometric modality with k-fold cross-validation. The score obtained by KNN classifier is used to verify the performance of the recognition system using the receiver operating characteristic (ROC) curve between the genuine acceptance rate (GAR) and false acceptance rate (FAR). The identification results of MFCC, GFCC and fused GFCC-MFCC at different noisy conditions are tabulated in Tables 1, 2, 3 and 4.

Table 1 Identification results of MFCC, GFCC, and fused GFCC-MFCC (F-GFCC-MFCC) at clean speech

Modality	False acceptance rate (FAR %)		Identification results (%)
	0.1	1	
MFCC	85	91.85	94
GFCC	86.1	92	95.2
F-GFCC-MFCC	90.1	94.6	98.2

Table 2 Identification results of MFCC at different noise with different SNR

SNR (dB)	Identification results (%)		
	Babble noise	Volvo noise	Destroyer engine noise
−5	51	43	7
5	75	67	15
10	79	71	69
20	83	74	70

Table 3 Identification results of GFCC at different noise with different SNR

SNR (dB)	Identification results (%)		
	Babble noise	Volvo noise	Destroyer engine noise
−5	64	51	29
5	83	78	34
10	87	83	77
20	91	84	81

Table 4 Identification results of fused GFCC-MFCC at different noise with different SNR

SNR (dB)	Identification results (%)		
	Babble noise	Volvo noise	Destroyer engine noise
−5	71.2	54.4	31
5	86.1	81.2	37
10	89.7	88.5	79
20	93.2	87.8	84.4

As shown in Fig. 4, in ROC curves of MFCC alone with 10 dB babble noise, GAR varies from 48 to 59 % at 0.1–1.0 FAR, respectively, and for GFCC, GAR varies from 57 to 71 % at 0.1–1.0 FAR, respectively. This shows that GFCC outperforms MFCC in the presence of noise. At score level fusion, for fused GFCC-MFCC (F-GFCC-MFCC), using sum rule, GAR is 91.93 % at 0.1 FAR, while at 1.0 FAR, GAR is 95.7 %. Similarly for product rule, GAR is 92 % for 0.1 FAR and 97.25 % for 1.0 FAR. For Hamacher T-norm, GAR is 78.36 % at 0.1 FAR, while at 1.0 FAR, GAR is 87 %. For Frank T-norm, GAR is 97.76 % at 0.1 FAR, while at 1.0 FAR, GAR is 98.8 %. As it is seen in the plots the ROC curve of score level fusion converges more rapidly as compared to the individual modalities showing the improvement in the performance of multimodal biometric system over unimodal biometric system.

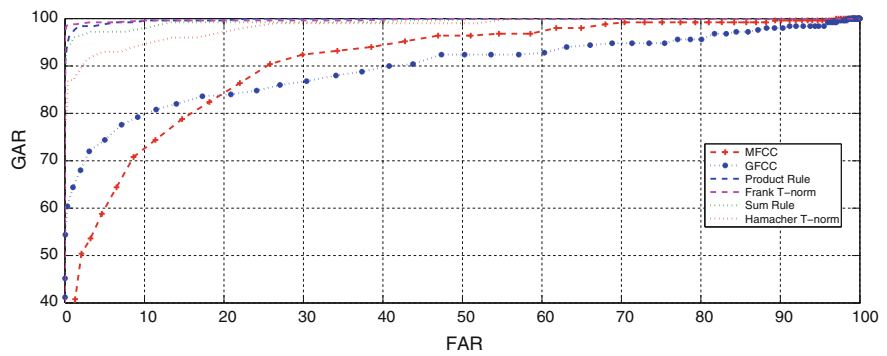


Fig. 4 ROC of MFCC, GFCC, and fused GFCC-MFCC (F-GFCC-MFCC) mixing babble noise at 10 dB

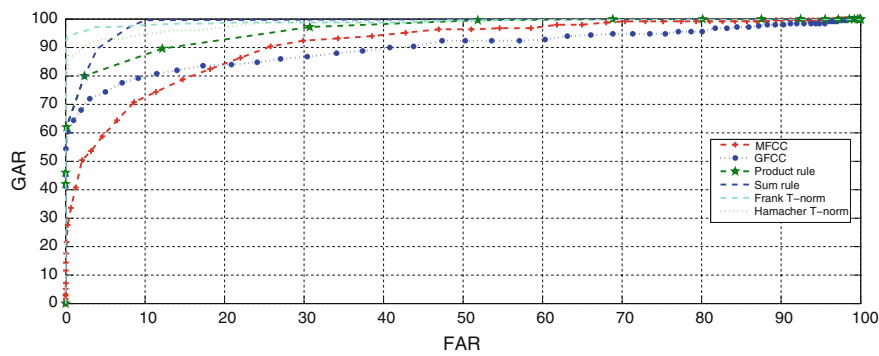


Fig. 5 ROC of MFCC, GFCC, and fused GFCC-MFCC (F-GFCC-MFCC) mixing volvo noise at 10 dB

At 10 dB Volvo noise, fused GFCC-MFCC (F-GFCC-MFCC), as shown in Fig. 5 using sum rule, GAR is 61.3 % at 0.1 FAR, while at 1.0 FAR, GAR is 69.6 %. Similarly for product rule, GAR is 61.3 % for 0.1 FAR and 69.6 % for 1.0 FAR. For Hamacher T-norm, GAR is 58 % at 0.1 FAR, while at 1.0 FAR, GAR is 66 %. For Frank T-norm, GAR is 92.1 % at 0.1 FAR, while at 1.0 FAR, GAR is 94 %.

At 10 dB destroyer engine noise, fused GFCC-MFCC (F-GFCC-MFCC), as shown in Fig. 6 using sum rule, GAR is 54 % at 0.1 FAR, while at 1.0 FAR, GAR is 64.6 %. Similarly for product rule, GAR is 61.9 % for 0.1 FAR and 71 % for 1.0 FAR. For Hamacher T-norm, GAR is 78.25 % at 0.1 FAR, while at 1.0 FAR, GAR is 87 %. For Frank T-norm, GAR is 62.1 % at 0.1 FAR, while at 1.0 FAR, GAR is 71 %.

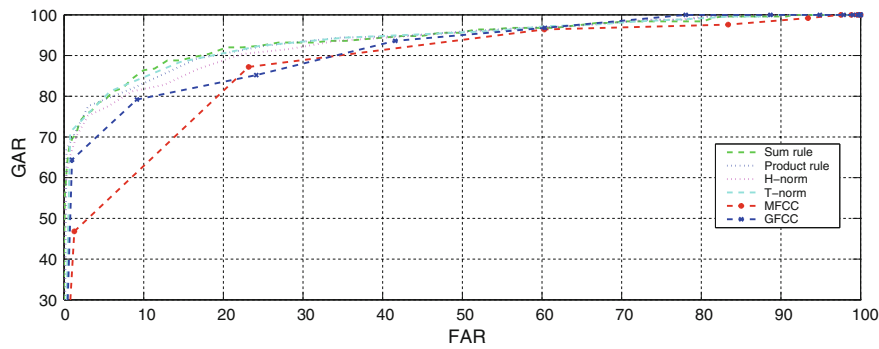


Fig. 6 ROC of MFCC, GFCC, and fused GFCC-MFCC (F-GFCC-MFCC) mixing destroyer engine noise at 10 dB

6 Conclusion

In multimodal biometric system, complementary information is fused to overcome the drawbacks of the unimodal biometric systems. Score level fusion of MFCC and GFCC using GMF-based feature extraction gives better performance over singular modalities. The results shown prove that the performance of multimodal biometric system is significantly improved as compared to the unimodal biometric systems.

References

1. Bhardwaj, S., Srivastava, S., Hanmandlu, M., Gupta, J.: Gfm-based methods for speaker identification. *Cybernetics, IEEE Transactions on* 43(3), 1047–1058 (2013).
2. Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: *Biometric Authentication*, pp. 731–738. Springer (2004).
3. Rodrigues, R.N., Ling, L.L., Govindaraju, V.: Robustness of multimodal biometric fusion methods against spoof attacks. *Journal of Visual Languages & Computing* 20(3), 169–179 (2009).
4. Wang, K.Q., Khisa, A.S., Wu, X.Q., Zhao, Q.S.: Finger vein recognition using lbp variance with global matching. In: *Wavelet Analysis and Pattern Recognition (ICWAPR)*, 2012 International Conference on. pp. 196–201. IEEE (2012).
5. Zhao, X., Shao, Y., Wang, D.: Casa-based robust speaker identification. *Audio, Speech, and Language Processing, IEEE Transactions on* 20(5), 1608–1616 (2012).
6. Shao, D., Rappel, W.J., Levine, H.: Computational model for cell morphodynamics. *Physical review letters* 105(10), 108–104 (2010).
7. Arora, P., Hanmandlu, M., Srivastava, S.: Gait based authentication using gait information image features. *Pattern Recognition Letters* (2015).
8. Arora, P., Srivastava, S.: Gait recognition using gait gaussian image. In: *Signal Processing and Integrated Networks 2015 (SPIN 21015)*, 2nd International Conference on. pp. 915–918. IEEE (2015).
9. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. *Pattern recognition* 38(12), 2270–2285 (2005).
10. Varga, A., Steeneken, H.J.: Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech communication* 12(3), 247–251 (1993).

Local and Global Color Histogram Feature for Color Content-Based Image Retrieval System

Jyoti Narwade and Binod Kumar

Abstract Content-based image retrieval system nowadays use color histogram as a common color descriptor. We consider color as one of the important features during image representation process. Different transformations such as changing scale of image, rotating an image, and translations of image to other forms does not make any alterations to the color content of image. If we need to focus on differentiation or similarity between two images we usually deal with various color features of image. To extract color features of image we consider on color space, color reduction, color feature extraction process. In image retrieval applications, user specifies desired image as query image and wants to search for the most similar image in database of his interest. Application then identifies similar relevant images from database based on different color features of database images and query image. To achieve this we compute color features of database images and those for query image. We use local color features of different regions and combine them to represent color histogram as a color feature. These color features are compared using Euclidean distance as a metric to define similarity between the query image and the database images. For calculations of local color histogram we divide image into different blocks of size 8×8 as fixed, so that for each block of image spatial color feature histogram of image is obtained. Our experimental work shows that local hybrid color histogram produced more accurate image retrieval results than global color moments color histogram.

Keywords Color space dimension reduction • Feature vector quantization • Low level color feature histogram • Global and local region color distribution • Ring-shaped concentric histogram and cornered histogram

Jyoti Narwade (✉)

Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India
e-mail: jyotinarwade@gmail.com

Jyoti Narwade

MAEER'S MIT College of Engineering, Pune, India

Binod Kumar

JSPM's Jayawant Technical Campus, Pune, India
e-mail: binod.istar.1970@gmail.com

1 Introduction

Content-based image retrieval systems aim for image indexing and retrieval in efficient and easier way so that actual human involvement during indexing process gets reduced [1]. To achieve this aim, a computer application has to be made capable enough to search and retrieve images from a database irrespective of specific manual annotations.

Content-based image retrieval system applications designed so far basically concentrate on considering each image as a complete to represent image feature. However, if we consider a single image, it has multiple areas called as subregions and objects. Each of them pertains some unique meaning. Many times user looks up for image with some specific region during retrieval process. In such situations, it becomes cumbersome and unhandy task if we follow entire image as feature. More suitable way is to consider image feature as set of regions [2]. These regions have different color, texture, and shape features as unique content key features and are used as inputs for content-based image retrieval process. Color content-based image retrieval system application aims at searching and retrieval of similar images from image databases based on color contributions within a query image. Different feature representations for color are color histogram, color moments, and color sets. These features are derived from image with easy mathematical calculations with an added advantage of distinct judgement in image retrieval.

Kasprzak et al. [3] have proposed global feature index. They aim to focus and analyze occurrences of each color in an image. They consider local feature index of an image as histogram which is calculated based on unique color representative. For this they split an image into subparts. For every subpart color representative is selected. It also depends on the type of image. They considered the type of image based on if it is a portrait image, event shot image, or scenery image. In case of event images, central region color plays important role than other subareas as such images have unique object in an image. Traditional color histogram does not consider region wise specific color distribution. It results in incorrect retrieval results due to incorrect similarity difference. To overcome such situations Fierro-Radilla [4] have proposed advances on color feature as color moments, color coherence vector, and color correlogram. Huang et al. [5] created different fuzzy regions color moments for central and computed color distribution for each region with respect to entire image. Quantization is applied to the reduce size of image. Sometimes it produces color information loss in neighboring pixels which directly has impact on overall image color feature. To overcome effects on image histogram, Shekar et al. [6] used color moments. They extracted color moments of all regions and applied region wise clustering. The mean value of each region moments is treated as one of the primary feature of the image. To define similarity among moments of two images Euclidean distance is calculated. Pass and Zabih [7] define color coherence vector as feature. To calculate it they split a color histogram into

two different parts: similar and dissimilar pixels called as coherent and incoherent pixels, respectively. Coherent pixel means pixels which have same color as with other pixels in image. It is like a region of pixels with same color otherwise it is incoherent pixel as it does not have same color value.

All the above work is based on considering various color frequencies in image. They did not consider the extent of how and where the color lies in an image. Hence such information remains unrecorded. In our work, we have implemented color spatial histograms to preserve distribution manner of colors at specific region and in particular direction. Furthermore instead of RGB color space, we focus on quantization and selection of uniform HSV color space so that appearance of colors remains unchanged due to quantization effect which is not taken into account in RGB color space.

2 Color Space Dimension Reduction

Image is made up of variety of different colors. To define similarity between two images there is a need of pixel to pixel mapping between images. Such pixel to pixel comparisons require a lot of computation time. As a result, it increases the running time of image comparison algorithms. To overcome this we start with quantization of image so that number of colors in feature vector is reduced. We select HSV as uniform color space for image feature representation to increase performance of image matching and retrieval process. Each axis is divided into equal-sized parts. Number of these parts is dependent on the scheme used for dividing the color space. For example divide the red and green axis into 8 segments each and the blue axis into 4 resulting in $8 \times 8 \times 4$ regions. Each of the original colors is mapped to the region where it falls in. Average of all colors getting mapped to particular region is considered as representative color for individual color [8].

3 Color Feature Vector Computation

Our system partitions an image into a number of homogenous regions and calculates local features for each region. These features of regions are used to represent entire image feature as global feature. Color feature does not get affected due to transformations such as rotation and scaling [9]. Hence it remains unchanged. This insensitive characteristic toward image transformations makes color feature as most important as compared with shape and texture features during image comparison process. Using histograms of local features along with reduction in dimensionality of color feature, we have tried to reduce image retrieval response time as number of color comparisons is reduced.

3.1 *Global Color Distribution Feature*

Each image in an image database can be different from remaining images but at the same time all images may share certain common characteristics. To preserve and monitor these common characteristics, we need the probability density distribution of various regions and the same to represent an image with some lesser numbers of bins. Color descriptions used in this paper are the mean value and the standard deviation of image. The mean value (μ) and the standard deviation (σ) of the color image are calculated as formulas proposed in [10].

3.2 *Local Color Distribution Feature*

During retrieval process we aim toward betterment of the retrieval accuracy. Hence, we used global features in first pass to filter few images from database. Then we use local features of those filtered images for further comparison with local features of query image. For this we used statistical information of color bins of image. To calculate region wise directional statistical histogram, we consider hybrid histograms which collect color occurrences count of image colors in specific direction to get directional locations, and color occurrences within specific central distance to get ring-shaped curved locations from a center point in each bin block [11].

3.2.1 *Curved Region Feature Vector*

As mentioned earlier, we quantized image in HSV color space into $8*8*4$ bin blocks. For every block in every color channel, we form ring-shaped curved locations with single center and different central edge length. Number of rings for each color can be varied from 1 to 8 depending on quantization level for better performance. Consider histogram subset S_q for each color bin B_q . Find centroid C_q as, $C_q = (X_q, Y_q)$. X_q, Y_q represents average sum of X and Y coordinates, respectively. Radius R when number of regions is only 1 is calculated using formula,

$$R = \sqrt{(X - X_q)^2 + (Y - Y_q)^2} \quad (1)$$

where X and Y are maximum coordinates. Curved distribution is a matrix $(|R_1|, |R_2|, \dots, |R_N|)$ if we formed N different ring structures. Curved regional color distribution feature is calculated by counting the number of points in each curved region. This process is repeated for all $8*8*4$ bin blocks.

3.2.2 Directional Regions

Directional regions are formed by dividing $8*8*4$ bin blocks into 8 cornered rings like regions called as quadrants with some fixed directional angle. We can decide number of regions depending on quantization level. For each color bin we calculate histogram subset S_q as count of points in each cornered region for each color bin B_q . To check if a point falls in a particular quadrant, we calculate its direction angle with respect to center point and positive X -axis. For each point $(X, Y) \in S_q$, calculate the direction $\Theta(X, Y)$ using formula,

$$(X, Y) = \arctan\left(\frac{y - y_q}{x - x_q}\right) \pm \Pi \quad (2)$$

where $+$, $-$ is selected depending upon in which quadrant point lies in. Then average direction $\theta(S_q)$ is called as principle direction [7] of S_q is obtained. Directional distribution is a matrix $(|R_1|, |R_2|, \dots, |R_N|)$ if we formed N different cornered ring structures. Directional distribution feature is calculated by counting the number of points in each cornered ring region. This process is repeated for all $8*8*4$ bin blocks.

4 Experimental Results

Following figures show results obtained at different stages of processing on image.

4.1 Color Space Conversion Outputs for RGB and HSV Image

See Figs. 1 and 2.

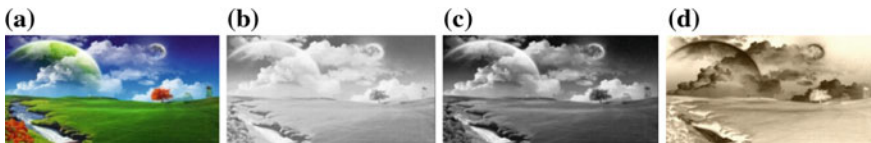


Fig. 1 a Original RGB image and b–d respective *red*, *green*, and *blue* color channels (color figure an online)

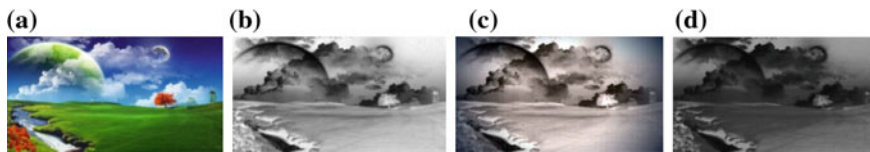


Fig. 2 **a** Original RGB image and **b–d** respective *hue*, *saturation*, and *value* color channels (color figure an online)

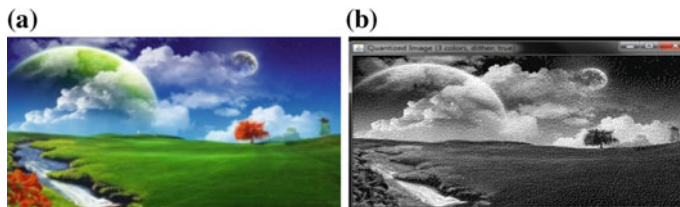


Fig. 3 **a** Input image, **b** three color-quantized image (color figure an online)

4.2 Quantized Image Outputs

Following figures show input image quantization using 3 colors and 5 colors. As an effect of quantization we achieved reduction in dimensions of color feature. Reduction is from $256 \times 256 \times 256$ dimension size to $8 \times 8 \times 8$ dimension size.

In Fig. 3a, b, we identified quantization difference by observing the dots obtained in image. These show the compactness of bin values of pixels having similar shade.

4.3 Similar Image Retrieval Results

User gives a query image to the color content-based image retrieval system. We provided graphical user interface for user to select query image. Irrespective of location of query image our image retrieval system computes hybrid color histogram and color moments as feature vector for query image which is compared with respective feature of every image stored in database. Ten most similar images are displayed in descending order of similarity difference.

4.3.1 Image Retrieval Results Using Local Hybrid Color Histogram and Global Color Moments Histogram

During retrieval, image which is having the smallest similarity difference is displayed as first image in sequence. Figure 4a shows hybrid color histogram for query image and Fig. 4b, c shows image retrieval result.

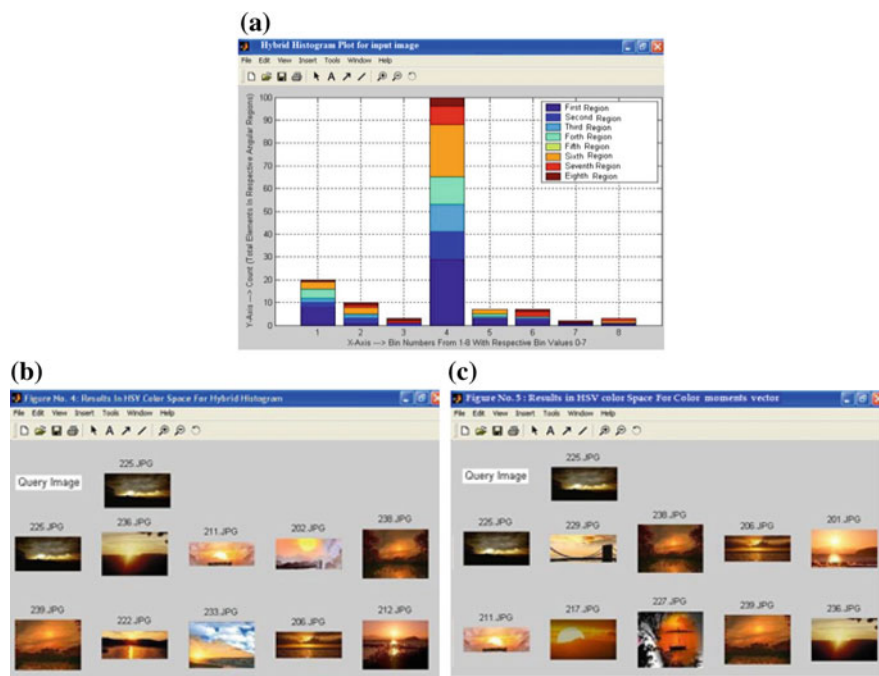


Fig. 4 **a** Hybrid histogram, **b** hybrid color retrieval, **c** color moments retrieval (color figure an online)

Above Fig. 4b, c shows 10 similar images retrieved using local hybrid color histogram and global color moments histogram from image database for sunset images.

5 Conclusion and Future Scope

By observing precision and recall values [12] for various retrieval results depending on query image, we found local hybrid color histogram produced more accurate results than global color moments color histogram. Our system works for similar color distribution at different regions. To achieve more accuracy in retrieval results, we can combine color feature extraction and object detection techniques so that location of object is taken into account.

References

1. Philipp Sandhaus, Mohammad Rabbath, Susanne Boll, "Employing Aesthetic Principles for Automatic Photobook layout" Springer Journal of Advances In Multimedia Modeling, 2011.
2. S. Orintara, T. T. Nguyen, "Using Phase And Magnitude Information of The Complex Directional Filter Bank for Texture Image Retrieval" IEEE International Conference on Image Processing, Volume 4, Pages 61–64, October 2007.
3. Włodzimierz Kasprzak, Wojciech Szykiewicz, Mikołaj Karolczak, "Global Color Image Features For Discrete Self-Localization of an Indoor Vehicle" Springer-CAIP 2005, LNCS 3691, Pages 620–627, 2005.
4. Atoany N. Fierro-Radilla, "An Efficient Color Descriptor Based on Global and Local Color Features for Image Retrieval" IEEE, International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), ISBN: 978-1-4799-1460-9, Pages 233–238, September 30-October 4 2013.
5. Z.C. Huang, P.P.K. Chan, W.W.Y. Ng, D.S. Yeung, "Content-Based Image Retrieval Using Color Moment and Gabor Texture Feature" In Proceedings of IEEE Ninth International Conference on Machine Learning and Cybernetics, Qingdao, pages 719–724, 2010.
6. B. H. Shekar, M. Sharmila Kumari, Raghuram Holla, "Shot Boundary Detection Algorithm Based on Color Texture Moments" Springer Communications In Computer and Information Science Volume 142, pages 591–594, 2011.
7. G. Pass, R. Zabih, "Histogram Refinement For Content-Based Image Retrieval" IEEE Workshop on Applications of Computer Vision, pages 96–102, December 1996.
8. Zhigang Xiang, "Color Image Quantization By Minimizing The Maximum Inter cluster Distance" ACM Transactions On Graphics, Volume 16, No. 3, July 1997.
9. P. S. Hiremath, Jagadeeshpujari, "Content Based Image Retrieval Based on Color, Texture and Shape Features Using Image and Its Complement" International Journal of Computer Science and Security, Volume 1, Issue 4, Pages 25–35, December 2007.
10. Tzu-Chuen Lua, Chin-Chen Chang, "Color Image Retrieval Technique Based on Color Features and Image Bitmap" Special Issue on AIRS2005: Information Retrieval Research n Asia, Information Processing & Management Volume 43, Issue 2, Pages 461–472, March 2007.
11. Jyoti Manoorkar, "Comparison of Spatial Color Histograms Using Quadratic Distance Measure" International Journal of Engineering Practices ISSN: 2277-9701, Volume 1, No. 1, April 2012.
12. Jyoti Narwade, Dr. M.M. Puri, "Comparative Study Of Spatial Color And Shape Features For Low Level Content Based Image Retrieval System" International Journal of Research in Computer Science And Management, IJRCSM, 2332–8088, July 2014.

Energy Efficient Pollution Monitoring System Using Deterministic Wireless Sensor Networks

Lokesh Agarwal, Gireesh Dixit, A.K. Jain, K.K. Pandey
and A. Khare

Abstract Wireless sensor networks (WSNs) are an organization of sensor nodes that interacts with each other remotely. Due to expansion of industrialization in the world, different types of pollution such as soil, air, radioactive increases day by day and it causes many health-related issues. In this paper, we conceive the problem of harmful gases CO₂, CO, SO₂, etc. We propose a comprehensive framework for detection and monitoring of air pollution using wireless sensor network. In the proposed framework, sensor nodes are deployed deterministically to cover region of interest with minimum number of nodes. Our system will monitor the real-time pollution with minimum delay. Better coverage with less number of nodes, minimum traffic from nodes to base station, balanced energy consumption are the main objectives of our proposed work.

Keywords Wireless sensor networks · Pollution monitoring · Deterministic deployment · Routing

1 Introduction

Wireless sensor networks (WSNs) are an organization of sensor nodes that interacts with each other remotely [1]. Sensor Nodes are spread randomly or manually over an area to check environmental or physical conditions depending on application.

Lokesh Agarwal · Gireesh Dixit
Madhav Prodyogiki Mahavidyalya, Bhopal, India

A.K. Jain (✉)
National Institute of Technology, Kurukshetra, India
e-mail: ankitjain@nitkr.ac.in

K.K. Pandey
Central Institute of Technology, Kokrajhar, India

A. Khare
UPES, Dehradun, India

Normally WSNs have a huge number of sensor motes and they have the capacity of intercommunicating with one another and they also communicate with a base station or sink.

Wireless sensor networks (WSNs) have been applied in numerous applications such as military surveillance, service monitoring, environment monitoring, etc. [2–4]. Sensor nodes have a restricted sensing and communicating range, so they are deployed in huge amount to cover the region of interest. The sensors coordinate among themselves to build a communication network such as one multi-hop network or a hierarchical system with several clusters and cluster heads [5]. Moreover, sensor networks works valuable role in the field of environment monitoring [6]. The main objective of this research is to design a real-time framework for detecting and monitoring the harmful substances in the air using wireless sensor network. The advantage of our framework is that, it can be deployed at any city or at any plant to monitor the pollution. If the values of harmful substances such as carbon dioxide, sulphur dioxide, etc., is greater than threshold level, then the node transmits information to base station with minimum time latency. Leakage of harmful gas from a industry is spread very fast, so reporting the leakage of gas to the controller with minimum time delay is also an important factor. In our framework, the sensor nodes deploy deterministically so it will cover the region with minimum nodes [7]. We have designed an efficient deterministic topology and routing strategy for this topology which ensures minimum collision between packets [8]. We have also designed cluster head selection algorithm which ensures that all the nodes become cluster head after some fixed time interval to increase the lifetime of system. All nodes within cluster send data to its respective cluster head by routing and then cluster head aggregates all the information and send fuse information to base station, moreover, in fusion process cluster head consumes very much of energy. If we do not change cluster head then it will die earlier and it affect the total lifetime of network. WSNs have a lot of unparalleled features which deliver many advantages and challenges in their application of real-time air pollution detection and monitoring system. Limited power resources of sensor network should be focused while building a framework for pollution detection system using WSNs.

To summarize the major contribution of paper is as follows:

- We design the real-time air pollution detection system, which can detect the leakage of harmful gas as early as possible.
- We propose a deterministic energy efficient cluster head selection protocol which increases the lifetime of sensor networks.
- We propose the framework which covers each and every point of region with minimum number of nodes.

The remainder of this paper is organized as follows: Sect. 2 of the paper presents some previous work and sensor use to detect the pollution. Section 3 presents the overview of the proposed approach and deterministic cluster head selection algorithm. Section 4 presents the implementation detail and results. We conclude the paper and present future scope in Sect. 5.

2 Related Work

There are two types of approaches used to monitor the pollutions, first is continuous online monitoring and another is passive sampling. In the passive sampling method, monitoring equipment is not providing the real-time values and only monitors the parameter at a certain time intervals. Sensor network is used for continuously monitoring system, furthermore, sensors sense environment parameter like concentration of gases from the environment and send this values by network to the centralized control centre. The way of data transfer contains both wired and wireless media. Wired method of data transmission is reliable and stable and with high speed of data communication. There are some limitations in wired mode of communication like expensive in installation, complex network, cabling, etc. Wireless mode in air and radioactive detection and monitoring system contains GPRS, GSM, WIMAX, etc. The advantages of WSNs is low cost, simplicity, liability and easy in installation. Our framework is grounded on the deployment of well-calibrated, comparatively low cost sensors node.

- A. *Nondispersive Infrared Sensor* Nondispersive Infrared Sensor (NDIR) is used to measure the concentration of gases like CO_2 , NO_2 , HO_2 , etc. [9]. It is a spectroscopic device used for gas detector [9]. NDIR sensor is used in many plants or in industry for measuring harmful gases. It is also used to assess the indoor air quantity. The instrument measures absorption of the individual wavelength of light when the gas diffuses or pumped into the light tube. Infrared sensor have sensitivities of 15–60 parts per million (PPM). NDIR carbon dioxide sensors are also applied in pharmaceutical fermentation, CO_2 sequestration and beverage carbonation applications as a dissolved carbon dioxide. Price of NDIR sensors are in the range of \$150–\$900 range.
- B. *Principle of NDIR sensor* The infrared light is sent by the sample chamber to detector and there is one more chamber with a surrounded reference gas [10], in this all light pass through the gas sample and only filtered before immediately before detector so it is called nondispersive (Fig. 1).

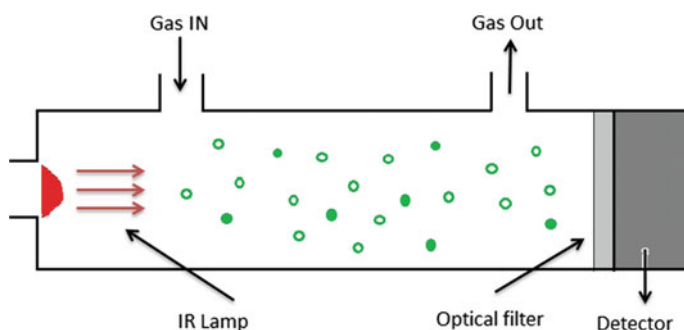


Fig. 1 Working of NDIR sensor

Detector has an optical filter, which absorbs only the wavelength of selected gas particles. To filter current background signals from the desired signal, the infrared signal from the source is normally chopped or modulated.

- C. *NDIR Sensor Design* Polychromatic light goes on by the gas sample and is engrossed in only selected wavelength. The gas sample disperses a metal trip in the top [10]. Reflection of light away the interval walls increases sensitivity. The filter ahead of the detector takes out all the light except that the corresponding to concentration of CO₂.

3 Proposed Framework

In this section, we present framework for detection and monitoring the air pollution using wireless sensor network. Our proposed framework contains the five parts: deployment of sensor node, clustering of nodes, intra-cluster routing protocol, inter cluster routing protocol and cluster-head selection algorithm. Sensor nodes are deployed in such a way that, it should cover each and every point of region. We deploy the nodes in triangular way, so that it covers the region with minimum number of nodes [11]. Afterwards, we make the clustering of nodes. Clustering saves the substantial amount of energy while transmitting the data and increase the lifetime of networks. Intra-cluster routing protocol defines a route between node and cluster head to transmit the packet. Inter-cluster routing protocol determines route from cluster head to base station. In the cluster head selection algorithm, cluster head is changed after a specific time interval, so that all nodes will become cluster head and increase the lifetime.

3.1 Deployment of Nodes

We have deployed sensor node deterministically in triangular fashion. The deployment of sensors is a key consideration as it affects the performance of system [11]. The goal of deterministic deployment is the following:

- Sensor nodes should deploy in a specific fashion to avoid the chances of collision and minimum transmission.
- In order to detect pollution timely, the sensor nodes should effectively cover the region of interest.
- Sensor node deploy such that it will cover whole region in minimum number of nodes and minimum routing between nodes (Fig. 2).

Fig. 2 Triangular grid deployment

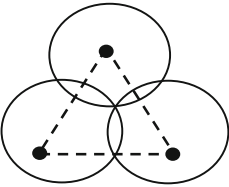
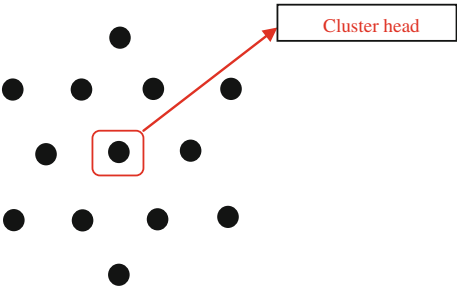


Fig. 3 Proposed cluster structure

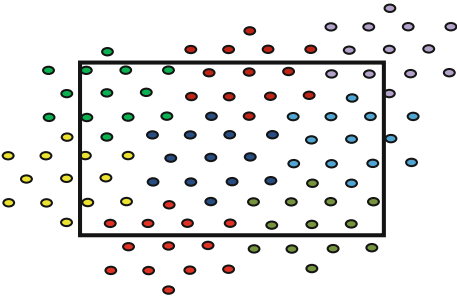


3.2 Clustering of Nodes

In our proposed framework, we construct the cluster of 13 nodes on the basis of distance. Cluster is constructed in such a way that, every node joins the cluster head of the nearest one. The structure of cluster is shown in Fig. 3.

Clustering reduces the power consumption of nodes and increase the total lifetime of sensor networks [12]. Head of cluster consumes more energy as compared to other nodes in the cluster, because it aggregates data from the other nodes and transmit to base station. Furthermore, we continuously change the position of cluster head and move the cluster in a deterministic manner so by this way we can enhance the lifetime of network [13]. In our proposed clustering, the one cluster which joins the cluster without overlapping is shown in Fig. 4. Nodes within a cluster can directly transmit data to its respected cluster head. After aggregating information, the cluster head transmits the data to base station or central controller. Afterwards a new node becomes the cluster head. The rightmost node of the

Fig. 4 One cluster is fixed in another cluster without overlapping



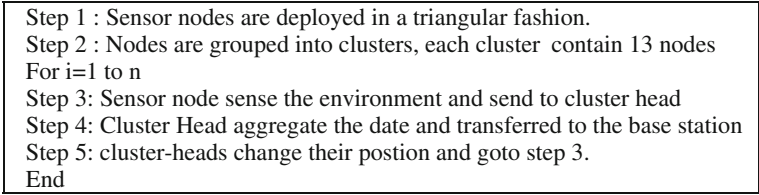


Fig. 5 Algorithm for choosing cluster head

previous cluster will become the new node and so on. The algorithm of clustering is shown in Fig. 5.

In this algorithm, it is ensured that all nodes will become cluster heads after 13 rounds when the total number of nodes is 169.

3.3 Inter-cluster Communication Protocol

It is the selection of path between cluster head to base station in efficient way. We choose the path on the basis of energy of nodes. We have taken the geometric mean of energy of all possible paths and choose the path having highest geometric mean. We prefer geometric mean rather than arithmetic mean because arithmetic mean includes the 0. For example, if path 1 has battery levels 6, 0, 3, 1 and path 2 have battery levels 3, 2, 3, 2. The arithmetic mean of both the paths is same but geometric mean does not include the path 1 because the geometric mean of path 1 is 0 and it does not include the dead node (Fig. 6).

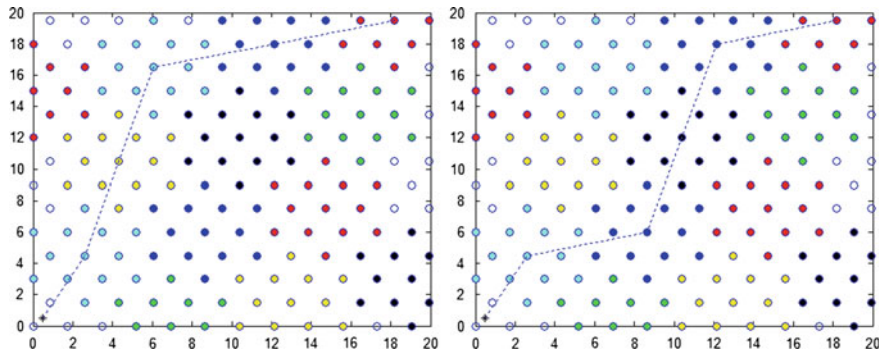


Fig. 6 Different paths from a node to base station

4 Implementation and Results

In this section, we will discuss the tools used for implementation and experiments results. We use MATLAB simulator to simulate our proposed framework. We deploy 169 sensor nodes in 12×14 dimension. We define the value of different parameters as shown in Table 1. We deploy the sensor nodes in deterministic manner in triangular grid as shown in figure. The vertical and horizontal coordinates of every sensor nodes must be preserved to calculate, the energy dissolution in data forwarding. Initially each node has 0.5 joule energy, in our simulation we are using this value but we can assign any value. It is necessary that all sensor nodes must have same energy before any action or event occurs. Sink has been placed corner of the field. Table 1 shows some detail about the energy consumption in numerous operations.

In this proposed deterministic triangular framework, we require less number of nodes as compared to square and hexagonal grid. The overlapping area is less as shown in Fig. 7.

We also simulate our framework using the square and hexagonal grid. When we increase the area of region then difference between nodes increases while keeping the sensing range same as shown in Fig. 8.

Lifetime of Network Now analysis the lifetime of this proposed network. We have done simulation with number of nodes 168. We have taken 3500 rounds in this simulation. After 3500 rounds there are 51 nodes are alive is showing in figure. First node died at round number 2431.

Table 1 Default values of parameter

Operations	Energy consumption
Electronic energy	50 nJ/bit
Amplifier energy(efs) (emp)	10 pJ/bit/m ²
	0.0013 pJ/bit/m ⁴
Energy for data aggregation	5 nJ/bit/signal

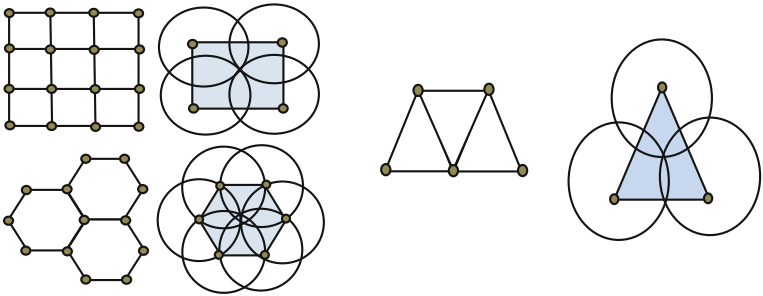
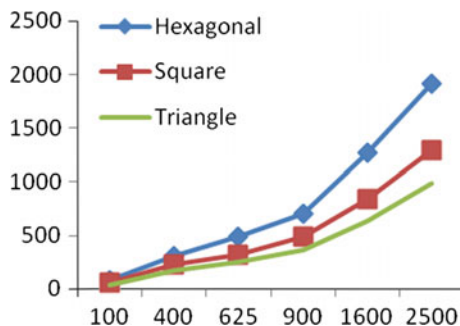


Fig. 7 Triangular, square and hexagonal grid deployment

Fig. 8 Comparison between different grids



5 Conclusion and Future Work

We present a comprehensive framework for detection and monitoring of air pollution which cover all parts of cycle of WSNs. We also propose an efficient clustering algorithm which guaranteed that every node will become cluster head. Efficient routing protocol chooses the path based on energy level. There may be several enhancements like routing between cluster head for better energy management. Furthermore, our framework presents significant data to the base station during the occurrence of harmful substance which will help in fixing the forecast procedure. Sensor nodes deployed in a hieratical fashion and a data aggregation method is applied for making fewer messages overhead during transmission. We can place node in some another fashion and design energy efficient and can apply a deterministic cluster head selection algorithm.

References

1. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Communications Magazine*, 40(8), 102–114 (2002).
2. Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D., Anderson, J.: Wireless sensor networks for habitat monitoring. In *ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02)*, Atlanta, GA, USA, (2002).
3. Bahrepour, M., Meratnia, N., Havinga P.J.M.: Fast and accurate residential fire detection using wireless sensor networks. *Environ Eng Manag J* 9(2):215–221 (2010).
4. Jain, A.K., Khare, A., Pandey, K.K.: Developing an efficient framework for real time monitoring of forest fire using wireless sensor network. *Second IEEE International Conference on Parallel, Distributed and Grid Computing*, Wanknaghat, Solan, Himachal Pradesh, India. 811–815 (2012).
5. O. Younis and S. Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, *IEEE Transactions on Mobile Computing* 3(4) 660–669 (2004).
6. Wenning, B.L., Pesch, D., Timm-Giel, A., Gorg, C.: Environmental monitoring aware routing: making environmental sensor networks more robust. *Telecomm Syst* 43(1–2), 3–11 (2010).
7. Mini, S., Udgata, S., Sabat, S.: Sensor deployment and scheduling for target coverage problem in wireless sensor networks. *IEEE Sens J* 14:636–644 (2014).

8. Younis, M., Senturk, I.F., Akkaya, K., Lee, S., Senel, F.: Topology management techniques for tolerating node failures in wireless sensor networks: A survey. *Computer Networks* (2013).
9. Peng, J., Ji, X-M.: A novel NDIR CO₂ sensor using a mid-IR hollow fiber as a gas cell, In proc. IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), 1489–1491 (2010).
10. Peng, J., Ji X-M, Shi YW, Liu Q., Bao ZM: A novel NDIR CO₂ sensor using a mid-IR hollow fiber as a gas cell. In proc. IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), 1489–1491 (2010).
11. Ma, M., Yang, Y.: Adaptive triangular deployment algorithm for unattended mobile sensor networks. *IEEE Transactions on Computers* 56, 946–947 (2007).
12. Jain, A.K., Purohit, N., Pandey, K.K., Jain, A., Bansal, H., Harshit, H.: An efficient clustering technique for deterministically deployed wireless sensor network. *Int. J. Comput. Application* 59(6), 975–8887 (2012).
13. Balamurugan, A., Purusothaman, T.: “IPSD, new coverage preserving and connectivity maintenance scheme for improving lifetime of wireless sensor networks.” *WSEAS Transactions on Communications*, 11, 744–761 (2012).

Development of Electronic Control System to Automatically Adjust Spray Output

Sachin Wandkar and Yogesh Chandra Bhatt

Abstract Agricultural pests and disease control is one of the main aspect of the crop production especially in orchards. Agricultural sprayers currently used in the orchards are basically constant rate air-assisted sprayers, which uses heavy stream of air to carry fine droplets toward the target. As, orchard trees vary in their size and geometry, spraying with these sprayers often results in either over spraying or under spraying. The variation in size and shape of the orchard trees makes necessary to take into consideration these tree parameters while spraying to adjust spray output to avoid losses due to over spraying. To adjust the amount of pesticide to be sprayed as per the tree size, an electronic control system was developed. The tree parameters were obtained using a high-speed ultrasonic sensor; a microcontroller was developed to control the system and to process the developed algorithm and variation in spray volume to be sprayed was achieved with a proportional solenoid valve. The developed system was tested in the laboratory to test the homogeneity of the flow rate sprayed through nozzle.

Keywords Pesticide application • Sensor • Variable rate spraying • Proportional valve

1 Introduction

In orchard production, protection of trees from pest and disease infestation is most important. In recent years, chemical control, i.e., foliar pesticide application has proved to be most effective pest and disease control method. To apply these pesticides, different equipment are used which uses high-pressure pump to pressurize

Sachin Wandkar (✉) • Y.C. Bhatt
Department of Farm Machinery and Power Engineering, College of Technology
and Engineering, MPUAT, Udaipur, Rajasthan, India
e-mail: Sachinwandkar85@gmail.com

Y.C. Bhatt
e-mail: drycbhatt@hotmail.com

liquid through nozzles to produce fine droplets. However, all these existing spraying equipment are constant rate applicators and spraying with these sprayers often results in over spraying. Unlike the field crops, orchard crops vary in their structure, size, and density. With the conventional sprayers, applicators are unable to manually adjust sprayer output to match the canopy size of the target trees, which results in wastage of pesticide. To control these losses, Morgon [12] stated the importance of taking into consideration the dimensions and other geometric parameters of the canopy to apply calculated amount of spray. This can be achieved using sensor technologies to identify trees and then apply precise amount of material needed for adequate pest and disease control.

Scientists have used ultrasonic sensors for canopy detection and characterization of plant canopy [7, 10, 15]. Using ultrasonic sensors, various researchers made efforts to develop an automatic spray control system with suitable proportional solenoid valve [1, 5, 8–11, 14]. Laser sensors were also found to be more suitable for detection of canopy occurrence, size and accordingly vary the spray output [2, 3, 13] but because of the low cost and simplicity, ultrasonic sensors have been extensively used.

The objective of this work was to develop an electronic control system using ultrasonic sensor to characterize the occurrence and width of canopy and then control the spray output to match the target tree structure.

2 Materials and Methods

The basic principle of the system was to control the output of the nozzle according to canopy characteristics and to shut the nozzle off when there is no canopy. The developed system consisted of a pesticide tank, pump, pressure regulator, and nozzle assembly together with an ultrasonic sensor, a microcontroller and a proportional solenoid valve (Fig. 1). The developed system was based on the electronic method for canopy width measurement and the amount of spray liquid was modified accordingly through proportional solenoid valve in order to achieve a proportional spray distribution based on crop geometry [4, 6].

2.1 Canopy Sensing and Characterization

A high-speed ultrasonic sensor with IP65 water and dust proof rating (Model UC 2000-30GM-1U-V1, Pepperl+Fuchs Group, Germany) was used for canopy detection and characterization. The power to the sensor was provided by a custom designed microcontroller board by regulating 12 V (DC) supply. The sensor was programmed to detect the objects in the range of 200–1700 mm. For the selected range, sensor was giving electronic output signal between 0 and 10 V. Laboratory

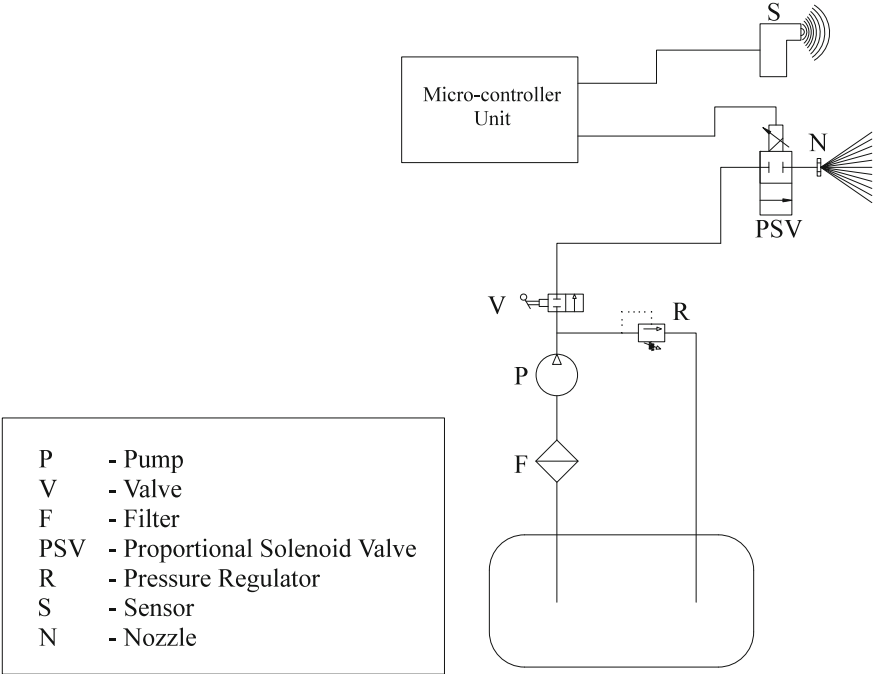


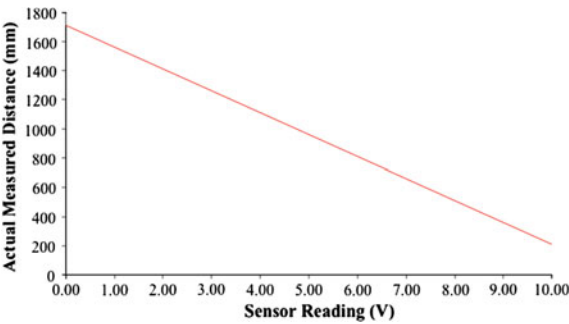
Fig. 1 Diagram of the elementary circuit with pump, pressure regulator, sensor, proportional solenoid valve, and nozzles

experiments were conducted to develop a relationship between measured distances and corresponding electronic output signal (Fig. 2). Following equation represents that relation.

$$d_i = -150.35v_i + 1712, \quad R^2 = 0.99 \tag{1}$$

where, d_i is the measured distance and v_i is the output electrical signal of the sensor.

Fig. 2 Correlation between actual distance measured and observed sensor reading



2.2 Microcontroller and Algorithm Design

The control algorithm was based on the measurement of canopy width (d). Once that parameter was electronically determined, information about system travel speed (v) and canopy height covered by single nozzle section (h) was added. The algorithm was developed in order to calculate the spray volume to be sprayed per unit time (Q), which was expressed in liter per minute. The main objective of the algorithm was to modify the delivered nozzle flow rate based on the measurements of the canopy volume along the crop line. Following equation represents the relationship applied for this process.

$$Q = d \cdot h \cdot v \cdot V_i \quad (2)$$

where, Q is the real time flow rate of individual nozzle (l/min.); v is the sprayer/tractor travel speed (m/min.); h is the height of canopy section covered by nozzle (m) (determined by dividing total canopy height with number of nozzle sections on one side of sprayer); V_i is the recommended application rate per unit tree volume (l/m³), and d is the depth of canopy (m).

In this study, recommended application rate (V_i) of 0.02741 l/m³ was selected by considering the tree row volume (TRV) of orchard crop and forward speed of 2.5 km/h (42 m/min) was considered. Depth (width) of canopy is calculated from canopy distance given by sensor (Eq. 3, Fig. 3) [6].

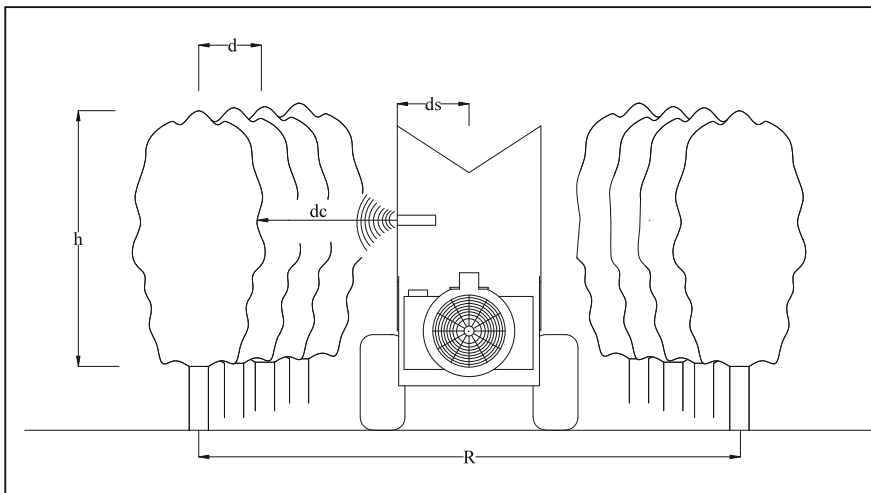


Fig. 3 Calculation of depth of canopy from the distance to the outside of the canopy (d_c) measured by the sensors: d_s , distance from sensor to sprayer axis; h , canopy height; R , tree row spacing

$$d = \frac{R}{2} - d_c - d_s \quad (3)$$

where, d is the depth of canopy (m); R is the tree row spacing (m); d_c is the distance between sensor and tree (m), and d_s is the distance between center of sprayer and sensor (m).

A custom designed microcontroller board (ATMEGA-328, ARDUINO) was fabricated to control the sensor and proportional solenoid valve. After the microcontroller was powered up, it triggered the sensor to begin acquiring the signals for canopy detection and width of canopy. The data acquisition system begins to receive information from the ultrasonic sensor. All data were then processed in the microcontroller, where signals acquired from the ultrasonic sensor were transformed first into width of canopy and then into intended flow rate, and finally into an electric control signal to be sent to the proportional solenoid valve. For each measured data, the system determines the distance from the sensor to the nearest tree foliage. According to Eq. 3, this value was transformed into crop width. All conversions were based on a defined orchard row-to-row spacing distance (R) and the assumption that the sprayer travelled along the centerline between rows [8]. Once the distance has been determined, the system transforms those values into the required flow rate according to Eq. 2 in order to apply the required amount of liquid in proportion to the orchard tree width variations.

The program flow chart for the microcontroller of the control system is shown in Fig. 4. The program started and stopped when the microcontroller power was turned on and off manually by the operator.

2.3 Proportional Solenoid Valve

To achieve the variation in flow rate as determined by the designed algorithm, the system was provided with a proportional solenoid valve. A normally closed proportional solenoid valve (model: Possiflow, ASCO Numatics, USA) with 1/4" size having maximum operating pressure of 8 bar was used. The valve was supplied with 24 V (DC) supply. The opening of the valve was controlled by the microcontroller by providing pulse width modulated control signal between 0 and 10 V according to size of the canopy. This pulse width modulation was achieved through external driver (L298, Motor and Solenoid Driver).

2.4 Testing of the Developed System in the Laboratory

To analyze the behavior of the selected proportional solenoid valve, the developed system was tested inside the laboratory conditions. During the test proportional solenoid valve was provided with known control signal between 0 and 10 V

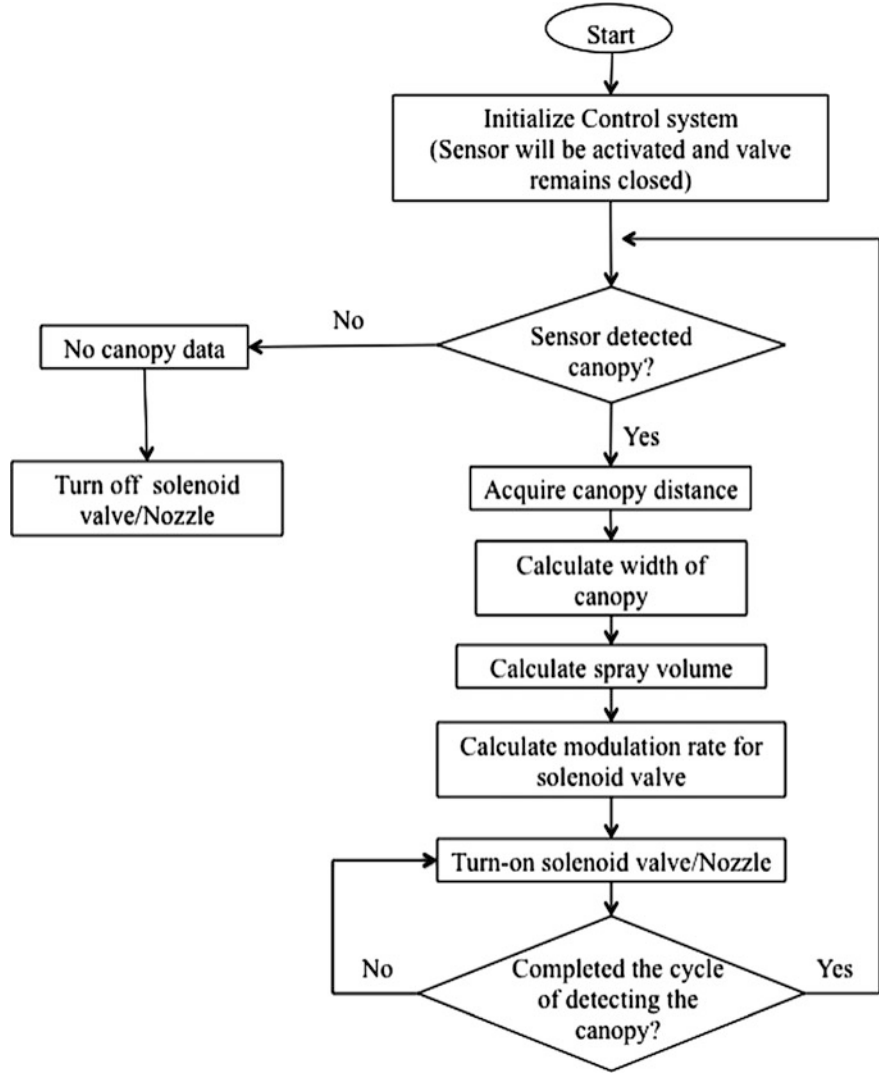
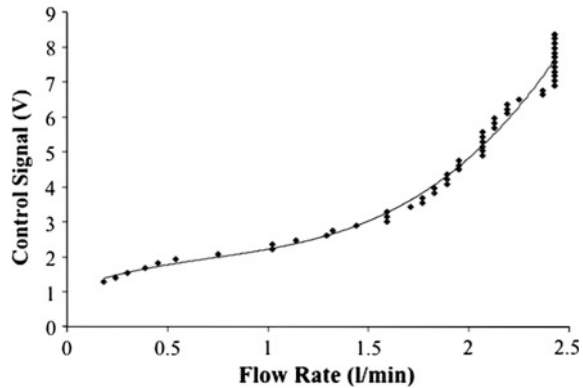


Fig. 4 Program flowchart for the microcontroller of the control system

corresponding to the size of the canopy and delivered flow rate was measured. The test was performed to establish the relationship between given control signal and corresponding flow rate delivered by the valve. During the testing, pump was operated at 3 kg/cm² pressure and it was kept constant throughout the supply system using suitable valves.

Fig. 5 Scatter diagram of the given control signal and delivered flow rate by the proportional solenoid valve



3 Results and Discussion

During the laboratory testing, per minute discharge of the valve was measured for provided control signal. Figure 5 shows the relationship between the flow rates delivered by the solenoid valve versus control signal provided. A cubic polynomial equation (Eq. 4) was developed which helps to determine the control signal to be sent to the solenoid valve to spray the determined flow rate. This equation is valid only within the selected range of parameters. The coefficient of determination for the developed equation was $R^2 = 0.98$.

$$V = 0.8445q^3 - 1.8131q^2 + 2.1341q + 1.0511 \quad (4)$$

where V is the control signal sent to the proportional solenoid valve (V) and q is the flow rate to be sprayed by an independent section (l/min).

4 Conclusions

An effort was made to develop an automatic control system to control the flow rate of the output nozzle according to the size of the tree to be sprayed. The system consisted of an ultrasonic sensor to automatically characterize the tree parameters. A suitable algorithm was developed to calculate the amount of flow rate (spray volume to be sprayed per unit time) needed to be sprayed through nozzle in accordance with the size of the tree in real time. The variation in flow rate was obtained with a proportional solenoid valve. To control the ultrasonic sensor and proportional solenoid valve, a customary microcontroller board was developed. The microcontroller processed the control algorithm after receiving signals from the sensor and subsequently controlled the proportional solenoid valve. The laboratory testing of the system produced relation between the signals to be sent to the valve to obtain required flow rate through nozzles.

References

1. Balsari, P., Doruchowski, G., Marucco, P., Tamagnone, M., Van de Zonde, J., Wenneker, M.: A system adjusting the spray application to the target characteristics. *Agricultural Engineering International: the CIGR E-journal*, 10 Manuscript ALNARP 08 002 (2008).
2. Chen, Y., Zhu, H., Ozkan, H.E.: Development of a variable rate sprayer with laser scanning sensor to synchronize spray outputs to tree structures. *Transactions of the ASABE*, 55 (3), 773–781 (2012).
3. Chen, Y., Ozkan, H.E., Zhu, H., Derksen, R.C., Krause, C.R.: Spray deposition inside tree canopies from a newly developed variable-rate air-assisted sprayer. *Transactions of the ASABE*, 56 (6), 1263–1272 (2013).
4. Escola, A., Rossel-Polo, J.R., Planas, S., Gil, E., Pomer, J., Camp, F., Llorens, J., Solanelles, F.: Variable rate sprayer. Part 1 – Orchard prototype: Design, implementation and validation. *Computers and Electronics in Agriculture*, 95, 122–135 (2013).
5. Gil, E., Escola, A., Rosell, J.R., Planas, S., Val, L.: Variable rate application of plant protection products in vineyard using ultrasonic sensors. *Crop Protection*, 26 (8), 1287–1297 (2007).
6. Gil, E., Llorens, J., Llop, J., Fabregas, X., Escola, A., Rossel-Polo, J.R.: Variable rate sprayer. Part 2 – Vineyard prototype: Design, implementation and validation. *Computers and Electronics in Agriculture*, 95, 136–150 (2013).
7. Giles, D.K., Delwiche, M.J., Dodd, R.B.: Electronic measurement of tree canopy volume. *Transactions of the ASAE*, 31 (1), 264–272 (1988).
8. Giles, D.K., Delwiche, M.J., Dodd, R.B.: Sprayer control by sensing orchards crop characteristics: Orchard architecture and spray liquid saving. *Journal of Agriculture Engineering Research*, 43 (4), 271–289 (1989).
9. Jeon, H.Y., Zhu, H.: Development of variable rate sprayer for nursery liner applications. *Transactions of the ASABE*, 55 (1), 303–312 (2012).
10. Llorens, J., Gil, E., Llop, J., Escola, A.: Variable rate dosing in precision viniculture: Use of electronic device to improve application efficiency. *Crop Protection*, 29 (3), 239–248 (2010).
11. Molto, E., Martin, B., Gutierrez, A.: Pesticide loss reduction by automatic adoption of spraying on globular trees. *Journal of Agricultural Engineering Research*, 78 (1), 35–41 (2001).
12. Morgan, N.G.: Gallons per acre of sprayed area: an alternative standard term for the spraying of plantation crops. *World Crops*, 16, 64–65 (1964).
13. Osterman, A., Godesa, T., Hocevar, M., Sirok, B., Stopar, M.: Real-time positioning algorithm for variable-geometry air-assisted sprayer. *Computers and Electronics in Agriculture*, 98, 175–182 (2013).
14. Solanelles, F., Escola, A., Planas, S., Rosell, J., Camp, F., Gracia, F.: An electronic control system for pesticide application proportional to the canopy width of tree crops. *Biosystems Engineering*, 95 (4), 473–481 (2006).
15. Tumbo, S.D., Salyani, M., Whitney, J.D., Wheaton, T.A., Miller, W.M.: Investigation of laser and ultrasonic ranging sensors for measurements of citrus canopy volume. *Applied Engineering in Agriculture*, 18 (3), 367–372 (2002).

Social Impact Theory-Based Node Placement Strategy for Wireless Sensor Networks

Kavita Kumari, Shruti Mittal, Rishemjit Kaur, Ritesh Kumar, Inderdeep Kaur Aulakh and Amol P. Bhondekar

Abstract The network density, energy consumption, and connectivity are the most important design parameters for a self-organizing wireless sensor network. This paper presents a social impact theory-based multi-objective strategy for optimizing these parameters. The proposed strategy optimizes the clustering schemes and signal strengths along with the operational modes of the sensor nodes. The algorithm has been implemented in MATLAB using an open source social impact theory Optimization toolbox (<http://mloss.org/software/view/457/>). The suggested algorithm offers the achievement of optimal designs and satisfies the different design parameters.

Keywords Component social impact theory • Network configuration • Sensor placement • Wireless sensor networks

Kavita Kumari (✉) · Shruti Mittal · I.K. Aulakh
Information Technology, UIET, Panjab University, Chandigarh, India
e-mail: kavita06it16@gmail.com

Shruti Mittal
e-mail: ershruti989@gmail.com

I.K. Aulakh
e-mail: ikaulakh@yahoo.com

Rishemjit Kaur · Ritesh Kumar · A.P. Bhondekar
Agtronics, Central Scientific Instruments Organisation, Chandigarh, India
e-mail: rishemjit.kaur@csio.res.in

Ritesh Kumar
e-mail: riteshkr@csio.res.in

A.P. Bhondekar
e-mail: amol.bhondekar@gmail.com

1 Introduction

Wireless sensor networks (WSNs) have incited remarkable research interests due to their vast potential in sensors, electronics, and computational fields. They have been exploited for civil as well as defense-related purposes. A WSN typically comprises large numbers of sensor nodes, which are energy constrained with limited computational and communication capabilities [1, 2]. The deployment of WSN nodes is usually based upon its application and could be random or deterministic [3–5]. Random deployment is usually done in hostile scenarios such as battlefield or hazardous environments, whereas amiable scenarios call for the deterministic deployment. In general, WSNs are expected to provide access to information about the physical world, regardless of time and space, this vision poses significant challenges for WSNs. The pervasiveness of WSN's limits its centralized control and is not practical and calls for capabilities of scalability, self-organization, self-adaptation, and survivability [6].

Energy utilization is a major issue for a WSN as the energy resources are consumed during the operation of nodes. The replacement of batteries or their recharge may sometimes be infeasible. Energy efficiency and utilization of a WSN depends upon the temporal resolution of information being collected, routing strategies, node placements, etc. [7–9]. Another important issues to be taken care of in a WSN are the network lifetime and connectivity. Cluster-based architectures are generally employed, in which the nodes are arranged in their network. These networks communicate with their respective cluster head node. Thus, collected information from the nodes is transmitted to the base station. The network connectivity problems include not only the load handling capability of the sink nodes, but also the ability of the sensor nodes to communicate with the cluster heads. Apart from the above issues, the application-specific design parameters also pose some issues. Several algorithms [3, 10–21] have been reported for the WSN design optimization in terms of scalability, self-organization, self-adaptation, and survivability. However, most of those suggested algorithms do not necessarily address the application-specific issues and make design parameterization and optimization a challenging task.

The design of a WSN system hence calls for simultaneous optimization of multiple nonlinear design parameters. This is a challenging task, as it requires finding pareto-optimal solutions under severe computational limitations. Such problems have been reported to be tackled with the application of computational approaches, such as neural networks, swarm optimization, genetic algorithm (GA), and ant colony optimization [22–28]. Social impact theory (SITO) is a recently introduced approach based on the application of a novel [29]. In this approach, a spatially distributed population of individuals in a two-dimensional lattice networks with each other to generate an optimal solution. In the process, the individuals change their attitude for a particular feature under influence of their neighbors' number, attitude, strength, and immediacy. The optimizer has been tested on benchmark problems for feature subset selection [29–33]. However, this optimizer has not been attempted for WSN optimization as yet.

In the present work, we have tried to analyze the application of SITO in WSN by integrating the network characteristics according to the application-specific requirements. In general, the algorithm under the constraints of application-specific requirements and energy consumption determines operational modes of the nodes. In particular, the network design has been investigated with respect to the sensor placements, communication range, and clustering. The performance of the proposed approach has been investigated by the study of connectivity and related energy characteristics and application-oriented properties (e.g., uniformity/spatial density of the sensing nodes). The work finally proposes an optimal design in which the mode of operation has been specified for each sensor node.

2 Methodology

2.1 Social Impact Theory-Based Optimization (SITO)

The social impact theory was proposed by Latané [34] wherein the author defined the social impact as any influence on an individual's feelings, thoughts, or behavior that is exerted by the real, implied, or imagined presence or actions of others. This meta-theory characterized the spatiotemporal variabilities of human opinion formation. This theory was modified by Nowak et al. [35] by taking into consideration the reciprocal influence of the individuals on their environment. Further, Macaš et al. [29] and Bhondekar et al. [30] implemented the above theory for optimal feature extraction and classification. The SITO algorithm is advantageous because of the requirement of few control parameters and capability of analyzing spatially distributed population.

In the SITO algorithm, an individual represents a probable solution for the problem at hand by maintaining a set of spatially distributed population in a two-dimensional lattice. The strength of the individual is estimated by taking into account the fitness value of its opinion. This opinion is subsequently modified at every iteration with respect to number of neighbors, strength, and immediacy. Total societal impact (I) is calculated by difference between the persuasive impact (I_p) of individuals holding the opposite opinions and the supportive impact (I_s) of individuals with the same opinion. I_p and I_s are defined as expressed by the following equations.

$$I_p = N_o^{1/2} \left[\sum (p_i/d_i^2) / N_o \right] \quad (1)$$

$$I_s = N_s^{1/2} \left[\sum (s_i/d_i^2) / N_s \right] \quad (2)$$

where, p_i is the persuasiveness of source i

s_i denotes the supportiveness of source i

N_o represents number of sources (individuals with opposing opinion)

N_s represents the number of individuals with individual opinions and d_i refers to the distance between the source i and the recipient

Generally, the individuals' opinions are modulated by comparing I_p and I_s , if I_p is greater than I_s , the of the individual changes with a probability $1 - K$. Similarly, the attitude may change with a probability K if I_p is lesser than I_s . The probability K improves the explorative capability by preventing loss of diversity.

The pseudocode as proposed by Macaš [29] is expressed as under:

```

Initialize attitudes by random assignment of binary values from
(0,1) to society.attitudes;
iter = 0;
WHILE (iter < max_iter) DO
    Compute society.fitness using Eq. 3 for corresponding
    society.attitudes;
    Find maximum fitness value, fmax, from society.fitness;
    Find minimum fitness value, fmin, from society.fitness
    Calculate society.strength = ( fmax - society.fitness ) / (
fmax - fmin );
    iter = iter+1;
    FOR each individual i and each dimension z DO
        Find sources and supporters in neighbourhood of i;
        Compute number of sources and supporters (No, Ns) in
        neighbourhood of individual i with respect to
        dimension z.
        Compute total persuasive impact  $I_p$  using Eq.1
        Compute total supportive impact  $I_s$  using Eq. 2
        IF ( $I_p > I_s$ ) and (i is not the best of all),
            Invert the attitude of individual i in
            dimension z with probability  $1-K$ 
        ELSE,
            Invert the attitude of individual i in dimension
            z with probability  $K$ ;
        END (IF)
    END (FOR)
END (WHILE)

```

2.1.1 Problem Outline

In this work we have assumed a two-dimensional field employing three types of sensors, which monitor parameters related to X, Y, Z. The spatial variability is such that sensor nodes' density in Z is greater than Y and X and for Y it is greater than X [36].

2.1.2 Network Model

A grid-based Euclidian model has been considered here, wherein the nodes have been placed at the intersections (see Fig. 1).

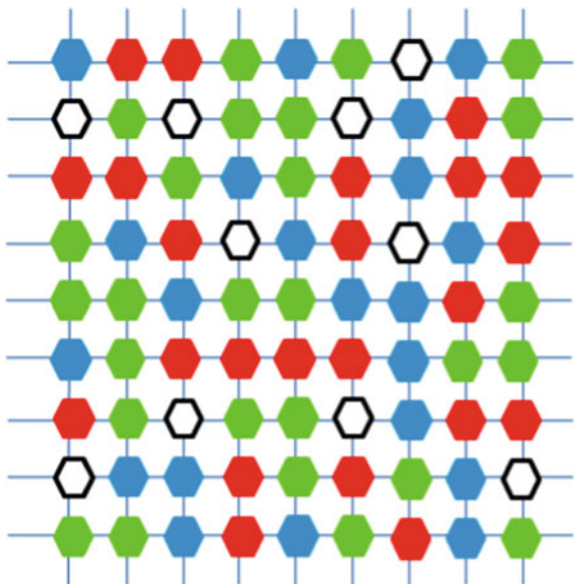
The active sensing nodes considered for this simulation are identical and have the usual features like power control and selection parameter for different sensing modes in X, Y, Z along with power control in transmission. We have assumed a cluster architecture where, the cluster-in-charge are the nodes operating in X-sense, along with Y and Z sensing modes with middle and small transmission ranges, respectively. It should be noted that the nodes present in the X mode can communicate with the base station using a multi-hop protocol and this leads to clustering of nodes in their vicinity [30].

Apart from sensing the X parameter the node in X-sense mode also performs tasks of data collection and its accumulation along with complex computations.

2.1.3 Problem Statement

The design parameters of WSN can be categorized into 3 classes [37]. First category incorporates the parameters of sensor deployment, e.g., uniformity and coverage. The second category deals with the connectivity parameters in a manner that no node remains unconnected. The last category involves the variables or parameters responsible for the survivability of the network, such as operational energy. In the proposed work, we have explored a multi-objective algorithm to optimally select these design parameters by scalarizing them into a single fitness function as

Fig. 1 The layout of a wireless sensor network



given by Eq. 3. The design optimization has been achieved by minimizing constraints such as number of unconnected sensors, operational energy and number of overlapping cluster-in-charge. The parameters namely number of sensors for each cluster-in-charge and field coverage are maximized. The measurement of quality of each probable solution of the optimization problem is given by the objective function in form of a numerical figure [36].

$$f = \min \left\{ \sum_{i=1}^5 k_i P_i \right\} \tag{3}$$

where, k and P_i are the corresponding weight and optimization parameter, respectively [36].

2.2 WSN Representation, Optimization Parameters, and Fitness Function

A square field ($L \times L$ length units) has been subdivided into several grids of unit lengths. The nodes are arranged on the grids. A bit-string represents an individual attitude in the society, which is employed for the encoding of the sensor nodes in a row-wise pattern as depicted in Fig. 2. Two bits are needed for the encoding of four states of the sensing nodes, viz. X, Y, Z, and inactive. Thus, the total length of the bit-string is set to be $2 \cdot L2$.

The optimization parameters listed in Table 1 are derived from following network attributes

- n_x is the X Sensors (cluster-in-charge) in terms of numbers. Similarly,
- n_y Y Sensors
- n_z Z Sensors
- n_{OR} Out-of-Range Sensors

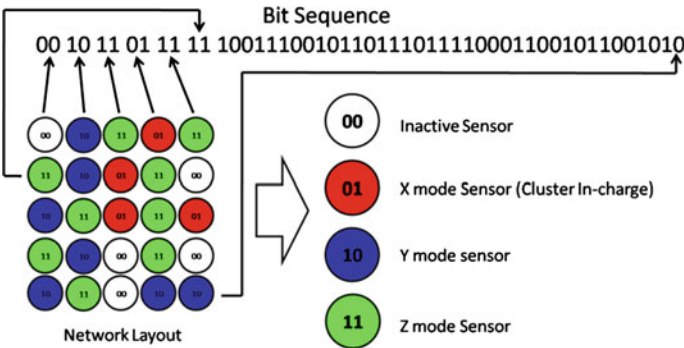


Fig. 2 Representation of Bit-string network layout [30]

Table 1 Correspondences between objectives and optimization parameters [36]

Objective	Parameters for optimization	Symbols
P1	Field coverage	FC
P2	Overlaps in unit cluster-in-charge error	OpCiE
P3	Out-of-range sensor error	SORE
P4	Sensors in unit cluster-in-charge	SpCi
P5	Network energy	NE

n_{inactive} Inactive Sensors
 n_{total} total sensing points
 n_o overlaps of cluster heads

The parameters to be optimized are derived as follows and are defined in [36]:

$$FC = \frac{(n_x + n_y + n_z) - (n_{OR} + n_{\text{inactive}})}{n_{\text{total}}} \quad (4)$$

$$SpCi = \frac{n_y + n_z - n_{OR}}{n_x} \quad (5)$$

$$SORE = \frac{n_{OR}}{n_{\text{total}} - n_{\text{inactive}}} \quad (6)$$

$$OpCiE = \frac{n_o}{n_x} \quad (7)$$

$$NE = \frac{4 \cdot n_x + 2 \cdot n_y + n_z}{n_{\text{total}}} \quad (8)$$

Therefore, there is a unique bit-string sequence for every unique WSN Design whose feature and performance can be estimated using fitness or weighting function. The fitness or weighting function needs to properly signify all the significant design parameters to influence the desired quality/performance of the WSN design. Each of the design parameter is equally important. Therefore for the present problem, the fitness function may be formulated as

$$f = -\alpha_1 FC + \alpha_2 OpCiE + \alpha_3 SORE - \alpha_4 SpCi + \alpha_5 NE \quad (9)$$

In the above fitness function, the appropriate weighting coefficients α_i ; $i = 1, 2 \dots 5$ define the significance of each design parameter. Therefore, the SITOs objective is to minimize the value of fitness function, to maximize some parameters their coefficients must be negative. The coefficient values are determined on the basis of design requirements and related experimentation. The desired values of the individual parameter coefficient were manually computed. The well-performing weight are listed in Table 2.

Table 2 Optimized values of the weighing coefficients

Parameters	Coefficient	Optimized value
Field coverage	α_1	6
Overlaps-per-cluster-in-charge error	α_2	0.65
Out-of-range sensors-error	α_3	9
Sensors-per-cluster-in-charge	α_4	1
Network energy	α_5	1.2

During the analysis (Table 2), the network connectivity variables (weights α_1 , α_4) were considered as constraints such that all the sensor nodes are in the limit of a cluster-in-charge, and no cluster-in-charge connects to higher than a predefined number of the sensors nodes.

3 Experimental Results

A network size of 10×10 was considered for experimentation. A total of 100 runs were carried out wherein four different machines performed 1000 iterations on four separate segments of 25 samples. Various combinations of society size and neighborhood were experimented and the best results in terms of convergence rate were obtained for society size of 225, and neighborhood size of 2. The convergence results obtained for neighborhood size 2 at different society sizes are shown in Fig. 3. It may be observed that increasing the society size beyond 225 does not

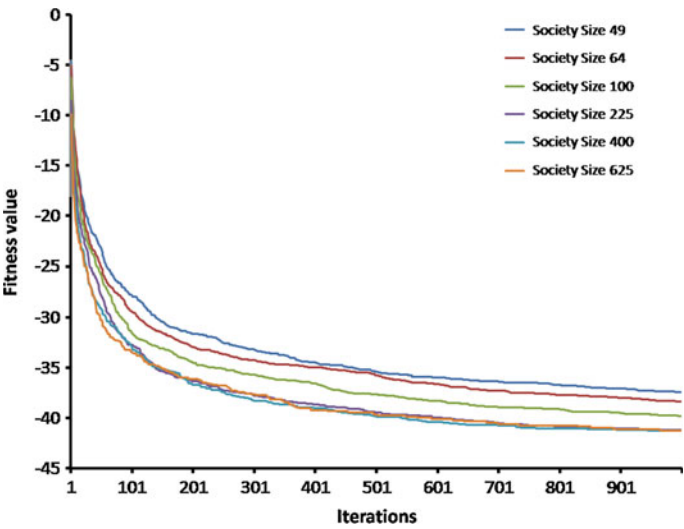


Fig. 3 A comparison between the convergence rates obtained at different society sizes with constant neighborhood size of 2

improve the performance of the algorithm. Moreover, by further increasing the society size, the convergence time increases exponentially. Similarly, Fig. 4 shows the convergence rate of the fitness value for a society size of 225 with varying neighborhood sizes. It may be observed that the neighborhood size of 2 gives the best convergence rate.

The optimized network by the algorithm is graphically represented employing a customized MATLAB script. One of the observed designs is shown in Fig. 5 in which the red, blue, and green circle, respectively, denote the X (cluster-in-charge),

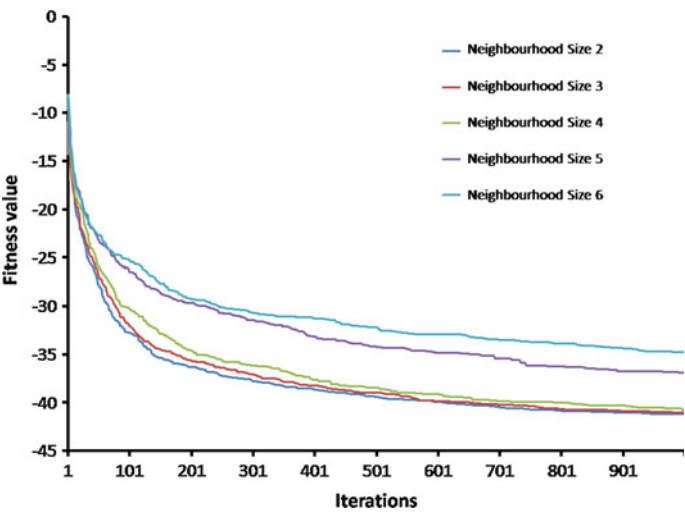
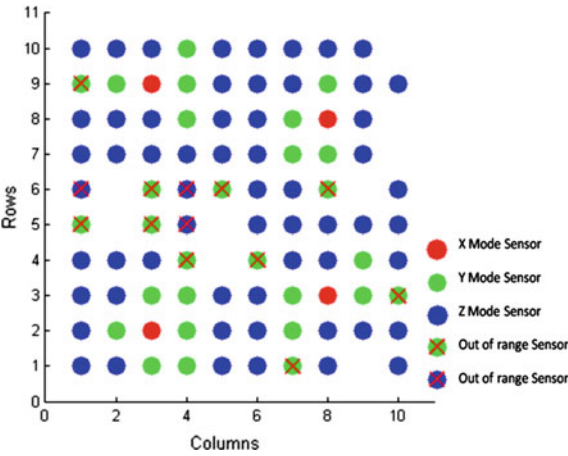


Fig. 4 Comparison of convergence rates obtained at different neighborhood sizes with constant society size of 225

Fig. 5 A graphically represented network as optimized by the algorithm



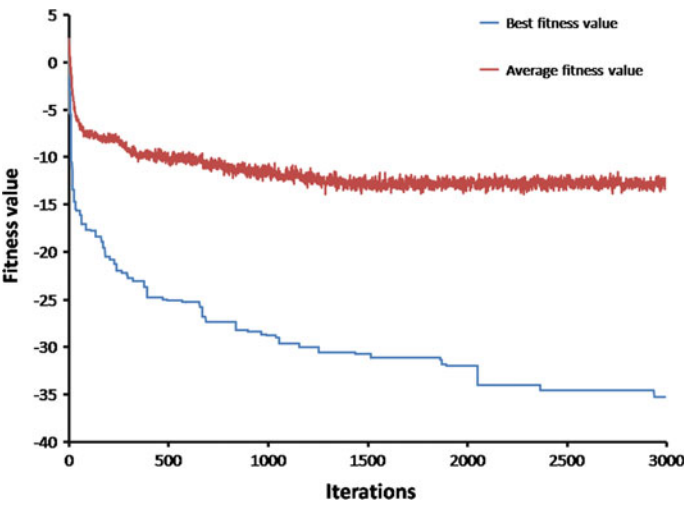


Fig. 6 Evolution progress for the identification of best individual (highest fitness value) and the entire society (average fitness value) using SITO approach

Y, and Z sensor positions. A cross-mark circle represents the out-of-range sensor node, whereas an inactive sensor node is represented by the empty space.

Figure 6 shows the progress of average value and minimum value of society fitness of one of the best runs. The three best SITO runs (abbreviated as S1, S2, and S3) that yielded the best results after 3000 iterations were observed and their results are in Table 3.

Table 3 Values of optimized parameters for 3 SITO-generated layouts of the network

Design parameter	S1	S2	S3
FC	0.85	0.9	0.75
OpCiE	2	1	0
SORE	0.2	0	0.01
SpCi	21.5	20.25	22.75
NE	1.6	1.5	1.41
Active sensors	91	87	93
X mode sensors	4	4	4
Y mode sensors	18	58	74
Z mode sensors	56	21	10
Inactive sensors	9	10	7
Out-of-range sensors	13	7	5
X mode or active sensors	0.043	0.044	0.043
Y mode or active sensors	0.197	0.644	0.795
Z mode or active sensors	0.615	0.233	0.107

4 Conclusions

A human opinion formation-based strategy (SITO) was used to optimize the nodes deployment of a fixed WSN. A grid-based fixed WSN having nodes of different operating modes was considered. The optimization was based upon various network parameters viz. field coverage, cluster overlapping, out-of-range errors, and network energy. The results showed that human opinion formation-based algorithm such as SITO can be used in WSN applications.

References

1. Yick, J., B. Mukherjee, and D. Ghosal, *Wireless sensor network survey*. Computer Networks, 2008. **52**(12): p. 2292–2330.
2. Rawat, P., et al., *Wireless sensor networks: a survey on recent developments and potential synergies*. The Journal of Supercomputing, 2014. **68**(1): p. 1–48.
3. Ishizuka, M. and M. Aida. *Performance study of node placement in sensor networks*. in *Distributed Computing Systems Workshops, 2004. Proceedings. 24th International Conference on*. 2004.
4. Izadi, D., J. Abawajy, and S. Ghanavati, *An Alternative Node Deployment Scheme for WSNs*. Sensors Journal, IEEE, 2015. **15**(2): p. 667–675.
5. Bhondekar, A.P., et al., *A multiobjective Fuzzy Inference System based deployment strategy for a distributed mobile sensor network*. Sensors & Transducers, 2010. **114**(3): p. 66.
6. Viani, F., et al. *Pervasive remote sensing through WSNs*. in *Antennas and Propagation (EUCAP), 2012 6th European Conference on*. 2012.
7. Kim, K.T., et al. *An energy efficient and optimal randomized clustering for wireless sensor networks*. in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015 16th IEEE/ACIS International Conference on*. 2015.
8. Shafali, et al. *An algorithm to minimize energy consumption using nature-inspired technique in wireless sensor network*. in *Circuit, Power and Computing Technologies (ICCPCT), 2015 International Conference on*. 2015.
9. Srbinovski, B., et al. *Energy aware adaptive sampling algorithm for energy harvesting wireless sensor networks*. in *Sensors Applications Symposium (SAS), 2015 IEEE*. 2015.
10. Slijepcevic, S. and M. Potkonjak. *Power efficient organization of wireless sensor networks*. in *Communications, 2001. ICC 2001. IEEE International Conference on*. 2001.
11. Krishnamachari, B. and F. Ordonez. *Analysis of energy-efficient, fair routing in wireless sensor networks through non-linear optimization*. in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*. 2003.
12. Zhou, C. and B. Krishnamachari. *Localized topology generation mechanisms for wireless sensor networks*. in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*. 2003.
13. Chen, S.Y. and Y.F. Li, *Automatic sensor placement for model-based robot vision*. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2004. **34**(1): p. 393–408.
14. Alageswaran, R., et al. *Design and implementation of dynamic sink node placement using Particle Swarm Optimization for life time maximization of WSN applications*. in *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*. 2012.
15. Deif, D.S. and Y. Gadallah, *Classification of Wireless Sensor Networks Deployment Techniques*. Communications Surveys & Tutorials, IEEE, 2014. **16**(2): p. 834–855.

16. Dondi, D., et al., *Modeling and Optimization of a Solar Energy Harvester System for Self-Powered Wireless Sensor Networks*. Industrial Electronics, IEEE Transactions on, 2008. **55**(7): p. 2759–2766.
17. Elshaikh, M., et al. *Energy consumption optimization with Ichi Taguchi method for Wireless Sensor Networks*. in *Electronic Design (ICED), 2014 2nd International Conference on*. 2014.
18. Hortos, W.S. *Bio-inspired, cross-layer protocol design for intrusion detection and identification in wireless sensor networks*. in *Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on*. 2012.
19. Kolega, E., V. Vescoukis, and D. Voutos. *Assessment of network simulators for real world WSNs in forest environments*. in *Networking, Sensing and Control (ICNSC), 2011 IEEE International Conference on*. 2011.
20. Menon, K.A.U., D. Maria, and H. Thirugnanam. *Power optimization strategies for wireless sensor networks in coal mines*. in *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on*. 2012.
21. Mohamaddoust, R., et al. *Designing the Lighting Control System Based on WSN with Optimization of Decision Making Algorithm*. in *Computational Intelligence and Communication Networks (CICN), 2010 International Conference on*. 2010.
22. Iram, R., et al. *Computational intelligence based optimization in wireless sensor network*. in *Information and Communication Technologies (ICICT), 2011 International Conference on*. 2011.
23. Lejiang, G., et al. *WSN Cluster Head Selection Algorithm Based on Neural Network*. in *Machine Vision and Human-Machine Interface (MVHI), 2010 International Conference on*. 2010.
24. Payal, A., C.S. Rai, and B.V.R. Reddy. *Artificial Neural Networks for developing localization framework in Wireless Sensor Networks*. in *Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on*. 2014.
25. Serpen, G., et al. *WSN-ANN: Parallel and distributed neurocomputing with wireless sensor networks*. in *Neural Networks (IJCNN), The 2013 International Joint Conference on*. 2013.
26. Singh, P. and S. Agrawal. *TDOA Based Node Localization in WSN Using Neural Networks*. in *Communication Systems and Network Technologies (CSNT), 2013 International Conference on*. 2013.
27. Subha, C.P., S. Malarkan, and K. Vaithinathan. *A survey on energy efficient neural network based clustering models in wireless sensor networks*. in *Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT), 2013 International Conference on*. 2013.
28. Wen-Tsai, S., et al. *Enhance the efficient of WSN data fusion by neural networks training process*. in *Computer Communication Control and Automation (3CA), 2010 International Symposium on*. 2010.
29. Macaš, M., L. Lhotská, and V. Křemen, *Social Impact based Approach to Feature Subset Selection*, in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2007)*, N. Krasnogor, et al., Editors. 2008, Springer Berlin Heidelberg. p. 239–248.
30. Bhondekar, A.P., et al., *A novel approach using Dynamic Social Impact Theory for optimization of impedance-Tongue (iTongue)*. Chemometrics and Intelligent Laboratory Systems, 2011. **109**(1): p. 65–76.
31. Kaur, R., et al., *Enhancing electronic nose performance: A novel feature selection approach using dynamic social impact theory and moving window time slicing for classification of Kangra orthodox black tea (Camellia sinensis (L.) O. Kuntze)*. Sensors and Actuators B: Chemical, 2012. **166–167**: p. 309–319.
32. Macaš, M., et al., *Binary social impact theory based optimization and its applications in pattern recognition*. Neurocomputing, 2014. **132**: p. 85–96.
33. Kaur, R., et al., *Human opinion dynamics: An inspiration to solve complex optimization problems*. Sci. Rep., 2013. **3**.
34. Latané, B., *The Psychology of Social Impact*. American Psychologist, 1981. **XXXVI** (4): p. 343–356.

35. Nowak, A., J. Szamrej, and B. Latané, *From private attitude to public opinion: a dynamic theory of social impact*. Psychological Review, 1990. **97**(3): p. 362–376.
36. Bhondekar, A.P., et al. *Genetic Algorithm Based Node Placement Methodology For Wireless Sensor Networks*. in *International MultiConference of Engineers and Computer Scientists*. 2009. Hong Kong: IAENG.
37. Ferentinos, K.P. and T.A. Tsiligiridis, *Adaptive design optimization of wireless sensor networks using genetic algorithms*. Computer Networks, 2007. **51**(4): p. 1031–1051.

Bang of Social Engineering in Social Networking Sites

Shilpi Sharma, J.S. Sodhi and Saksham Gulati

Abstract This research paper is a brief study on social engineering that explores the internet awareness among males and females of different age groups. In our study, we have researched on how an individual shares his/her identity and sensitive information which directly or indirectly affects them on social networking sites. This information can be user's personal identification traits, their photos, visited places, etc. The parameters chosen for influence of social engineering in social networking sites are passwords, share ability, and awareness. This research briefly explains how people between age group of 13–40 years share their information over the web and their awareness of netiquettes. This information is then conclusively used to calculate average amount of sensitive information which can be extracted through social engineering for different age groups of males and females.

Keywords Social networking site • Social engineering • Share ability • Awareness • Passwords • Victim

1 Introduction

In today's world where technology is the necessity in everybody's life, social engineering is emerging as vital vicinity in social networking sites. Different services are available for individuals, enterprises, and organizations that have implemented variety of features in social networking sites. These sites provide a perfect platform for hackers and attackers. Information posted and shared by users are always under threat.

Shilpi Sharma (✉) • Saksham Gulati
ASET, Amity University, Noida, India
e-mail: ssharma22@amity.edu

Saksham Gulati
e-mail: saksham.gulati28@gmail.com

J.S. Sodhi
AKC Data Systems Pvt. Ltd., Amity University, Noida, India
e-mail: jssodhi@amity.edu

Social Engineering, in information security refers to as psychological manipulation of human mind to extract sensitive information [1]. It is an art of deceiving the victim fetching sensitive information that can benefit the hacker or cracker [2, 3]. Social engineering is a widely used technique for extracting information from people to process, counter and plot a structured cyber attack. It is an approach that helps to crack the personal data of unknown users, to find their weaknesses or strengths for an organized crime. In 2011, a survey was conducted by Dimension Research in U.S, Canada, Australia, U.K, Germany, and New Zealand on IT professionals and concluded that 48 % are victims of social engineering attacks in social networking sites [4].

It has been found that the most significant security risks are associated with social engineering [5]. With the changing threat scenario in cyber space [6], hacking skills of hackers are becoming sophisticated and difficult to track. Social networking sites are progressively accessed by users of different age groups from teenagers to old people and the irony is users by pass their concern toward information security [7, 8]. Furthermore, information on social networking sites is accessed automatically by social engineering bots by providing data in machine readable form. The most common ways of social engineering includes distribution of adware's, uploading explicit content as advertisement, distribution of malware through ads, prank calls, surfing through web, fake emails, uploading of false information, etc.

Our paper explores the implications of age and gender in social engineering to fetch the password and know about the awareness of respondents, while sharing information in social networking sites. In our evaluation, we test our approach on gender and age of users on social networking sites using three basic measures, i.e., passwords, share ability, and awareness. As maturity comes with age and experiences in both the genders, our primary dataset is categorized into high, medium, and low level. It presents the variations in password of males and females of different age groups and awareness of sharing information over the internet [9, 10]. Social engineering is the biggest threat both at internal as well as external level for any company or an individual [11]. Thus, social engineering can be made easy by making them vulnerable to cyber crimes.

The rest of the paper is organized as follows: Sect. 2 summarizes research related to social engineering in social networking sites, Sect. 3 describes the methodology and result of concept implementation is outlined in Sect. 4. In Sect. 5, we draw conclusions from our findings and propose future research.

2 Related Work

Social engineering attacks are not only well known in practice but also in literature [12, 13]. Instead of pointing toward vulnerabilities in technical systems, the social engineering targets the weaknesses of people. Research on privacy implications of social networking sites has been discussed in a number of publications. The most

widely used social engineering techniques include social surfing, dumpster diving and shoulder surfing. These techniques are used by hackers on everyday basis to gather information about the victim [14]. Password guessing is a common way to crack passwords as no major risk is associated with it [15, 16]. Password guessing is mostly a psychological act where technology or softwares are not the primary factor [17]. The main motive of social engineering is to crack sensitive information of a victim as passwords related information holds the top priority [15–17].

Social engineering is totally dependent on an individual's personality [17]. A survey states that people with unstable personality can be manipulated easily for extracting information through them [15]. And people with strong personality do not share their sensitive information easily and mentioned social engineering as an internal threat [18]. Generally, individuals choose password based on their traits and if an attacker understands an individual thoroughly then sensitive information can be easily extracted [19]. Thus, it also provides the importance of training given to every user to prevent information against social engineering [20].

Social engineering comes as a message in the form of request that requires victims to accept or respond [21]. The attacker creates multiple fake profiles that impersonate with victims friend, relatives, or a famous person in social networking sites. Although many organizations control security threats but sometimes fails to recognize the dangers associated with social engineering attacks [7].

3 Research Methodology

The basic idea of research methodology states that “every mind can be tricked and manipulated” [4, 9]. The statement indicates that the most secure system in this world can be cracked through human hacking or social engineering [19]. For the study of awareness and sharing passwords, 400 samples with equal number of males and females are chosen and they are studied for a particular interval [23]. All their online social activities are recorded as a part of research to collect primary data. Their passwords were gathered and classified into three categories easy, medium, or difficult to crack [22]. To capture the potential personnel awareness and share ability, we include age and gender in the survey. Age and gender has been studied to come across the social engineering threats and effectiveness of internet security [23, 24].

4 Results

The sample size of our research was 400 which comprises of 200 males and 200 females to define their authenticity and security while creating passwords. This study is conducted on three social networking sites—Instagram, Facebook, and

LinkedIn. LinkedIn mostly have professional profiles hence high-quality data was shared while Instagram had profuse database of personal photographs. Facebook was significant to link the accounts with assured authenticity of data.

4.1 Password

Password is a basic login criterion to access any account in social networking sites. They are the most important credentials for logins and must be secured properly. Passwords must contain both upper and lower case characters along with special characters to make it strong. In our research, a study has been conducted among males and females of different age groups which have been categorized into two defined age groups with 13–20 years and 21–35 years.

The chart shows the level of difficulty, in terms of complexity of passwords:

- (1) Difficult: The passwords which have a combination of uppercase and lowercase alphabets along with special characters and have no usual meaning in any language are classified as difficult. Such passwords are difficult to crack, guess, or even shoulder surf [25].
- (2) Medium: These passwords generally contain both uppercase and lowercase alphabets with special characters. However, they can be easily guessed or cracked because they are either close to predefined dictionary word or have a meaning related to something that user generally talks about.
- (3) Easy: These passwords are very easy to guess as they generally have no mixing of uppercase and lowercase alphabets. Also, the passwords are short in length and carry a meaning closely related or associated to users.

Figure 1 shows that maximum users lying in male category have password of medium level. This interprets that password lacks combination of uppercase and lowercase characters along with special characters that makes it easy to crack using predefined dictionary. This category mostly comprise of teenagers who use internet on regular basis. As female use easy passwords, it concludes that the respondents are not concerned about the password leakage due to lack of knowledge about cyber crimes.

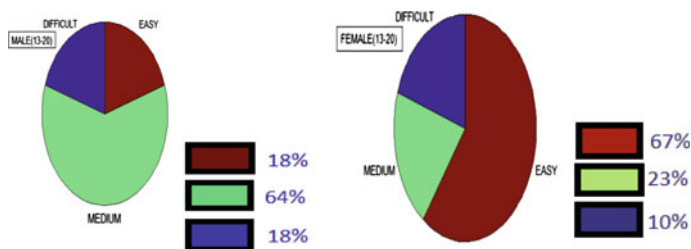


Fig. 1 Pie-chart showing male and female response for age group 13–20 years for password

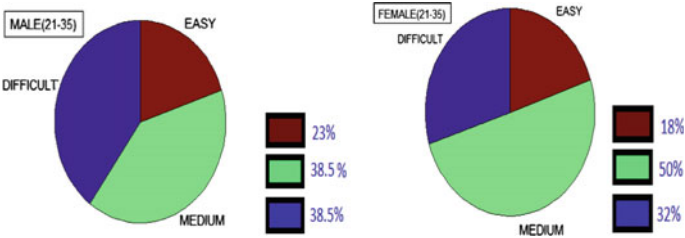


Fig. 2 Pie chart showing response of males and females of age 21–35 for password

Figure 2 explores a general change toward males and females in age group of 21–35 as they have shifted or inclined to a more secure password. The figure represents shifts from easy to difficult level of passwords. It can be clearly concluded from the result that age is proportional to the maturity of mind. Hence social engineering of a person in minor age group is easier.

Thus, the result shows that male from age group 21–35 creates more difficult passwords than females. And females store easy passwords that changes drastically as per age groups which develops with maturity and experience.

4.2 Share Ability

Data share ability is the measure of data that one shares on social networking sites through which a user can be classified as a potential victim to hacker.

Different levels of share ability with respect to age and gender are categorized as

- (1) High: Too much sensitive data is shared on social networking sites which can be used against the user that can be unsafe. This includes phone number, address, private photos, daily movements, etc.
- (2) Medium: The data or information shared by user is as per the requirement of social networking sites so, not much data is shared.
- (3) Low: People lying in this category share very less amount of data. Only a few pictures are uploaded with no personal information. People in this category generally do not show much interest in social networking sites.

Figure 3, represents the amount of data shared by males and females of different age groups.

Moreover, males and females of age group 21–35 show that they share very high amount of data on internet (Fig. 4).

Here we found that as per the demand of social networking site, people shares good amount of personal data. Males of age group 13–20 generally shares more information. Also, change in age group data directly influences share ability. With growing age, users gain experience of cyber world and cyber crimes that influence data share ability in both males and females.

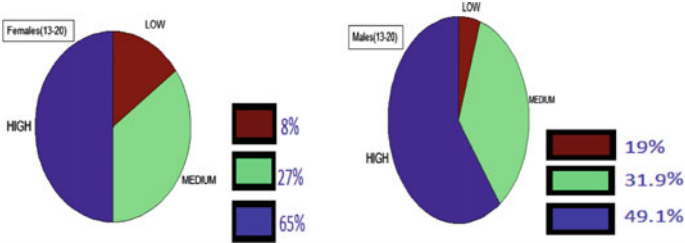


Fig. 3 Pie chart showing share ability of female and male of age group 13–21

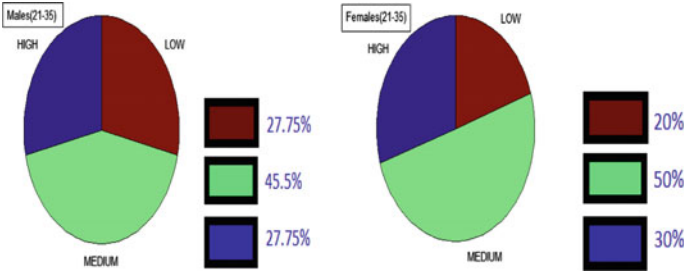


Fig. 4 Pie chart showing share ability of male and female of age group 21–35

4.3 Awareness

Awareness can be defined as a measure of knowledge about the crimes related to internet. A responsive user knows about the consequences of uploading sensitive information on social networking sites and can easily identify how to protect information leakage within the social networks. Also, user may share his photos and phone number on Facebook and restrict it to be viewed by few only.

Based on different levels of awareness with age and gender are categorized as:

- (1) High: This category generally contains technically sound engineers and professionals or prompt users of social networking site. People lying in this category share very high amount of sensitive information but they know how to protect it.
- (2) Medium: People of this category have some idea about private security on social networking site. They never use high profile security system like two-way authentication nor do they reply to requests over email to showcase their profile.
- (3) Low: People of this category do not have much idea about the usage of their sensitive information by providers or third parties of social networking sites. These persons generally send and accepts friend request to and by unknown people (Figs. 5 and 6).

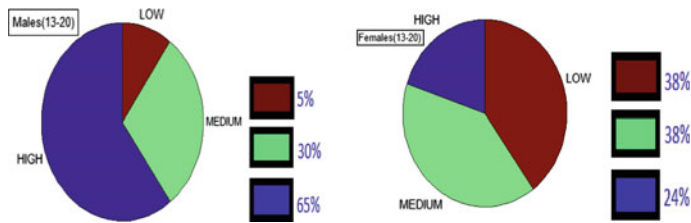


Fig. 5 Pie chart showing awareness of male and female of age group 13–20

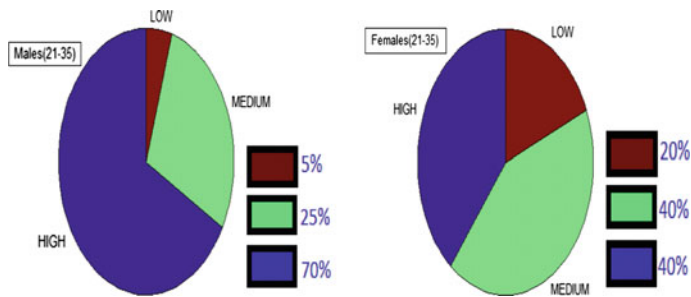


Fig. 6 Pie chart showing awareness of male and female of age group 21–35

The result shows that males are generally aware about the cyber crimes. A huge amount of data is shared over the internet and becomes a necessity to provide security otherwise the user will be trapped as a victim toward cyber crime. High awareness includes good knowledge of privacy over social networking sites that include hiding photos from anonymous, two token authentication and secondary email to reset passwords. It has been observed that most people in age group 21–35 of both genders had enough knowledge about cyber crimes. And females of age group 13–20 had limited knowledge about cyber crimes as compared to males of same age group.

So, it is pragmatic that males have high awareness among internet crimes and knows how to hide their private data or safeguard themselves from being a victim. A good variation was found in female category in terms of awareness with age.

5 Conclusions and Future Work

Social engineering can be used to exploit any human vulnerability either emotional or psychological and our study clearly shows that females are weaker as compared to the males. It was found during the study that females are emotionally weak as compared to males which often results in wrong decisions. Also females share more information and apply comparatively weaker passwords, hence it concludes that

females in general are easy target for social engineering. As deduced from the survey, it can also be concluded that males are more cautious than females of same or different age groups. They are more aware of the consequences of cyber crimes. With age comes the maturity and awareness about cyber experiences. So seniors keep track of their passwords.

Social engineering is a never ending threat to Information. Social engineering can only be prevented by means of experience since there is no formal professional training defined for the same. More of the training or knowledge is required for females of every age group. The user should be trained and made aware of social engineering threats inclusive of the factors that may cause serious attacks.

Practical training including psychology sessions are some of the best ways to train professionals against social engineering.

References

1. Irani, Danesh, Marco Balduzzi, Davide Balzarotti, Engin Kirda, and Calton Pu. "Reverse social engineering attacks on online social networks." In *Detection of intrusions and malware, and vulnerability assessment*, pp. 55–74, Springer Berlin Heidelberg (2011).
2. Mitnick, Kevin D., and William L. Simon.: *The art of deception: Controlling the human element of security*. John Wiley & Sons (2011).
3. Huber, Markus, Stewart Kowalski, Marcus Nohlberg, and Simon Tjoa.: *Towards automating social engineering using social networking sites*. In: *Computational Science and Engineering*, vol. 3, pp. 117–124. IEEE (2009).
4. Algarni, Abdulmohsen, Yue Xu, and Thomas Chan.: *Social Engineering in Social Networking Sites: The Art of Impersonation*. In: *Services Computing*, pp. 797–804. IEEE. (2014).
5. Hadnagy, Christopher.: *Social engineering: The art of human hacking*. John Wiley & Sons (2010).
6. Thakral A., Rakesh N. and Gupta A.: *Area Prone to Cyber Attacks.*, vol. 39, pp 40. CSI Communications (2015).
7. Sharma, S., & Sodhi., J. S.: *Social Network Analysis & Information Disclosure: A Case Study*. *International Journal of Computer, Electrical, Automation, Control and Information Engineering* vol: 9, pp. 567–575. WASET (2015).
8. Long, J. *No tech hacking: A guide to social engineering, dumpster diving, and shoulder surfing*. Syngress. (2011).
9. Graves, K. CEH: *Official Certified Ethical Hacker Review Guide: Exam 312–50*. John Wiley & Sons. (2007).
10. Uebelacker, S., & Quiel, S. *The social engineering personality framework*. In: *Socio-Technical Aspects in Security and Trust (STAST)*, pp. 24–30. IEEE. (2014).
11. Greitzer, Frank L., Jeremy R. Strozer, Sholom Cohen, Andrew P. Moore, David Mundie, and Jennifer Cowley.: *Analysis of Unintentional Insider Threats Deriving from Social Engineering Exploits*. In: *Security and Privacy Workshops (SPW)*, pp. 236–250. IEEE. (2014).
12. Stringhini, G., Kruegel, C., & Vigna, G.: *Detecting spammers on social networks*. In: *26th Annual Computer Security Applications Conference* pp. 1–9. ACM. (2010).
13. Webb, S., Caverlee, J., & Pu, C.: *Social Honeypots: Making Friends With A Spammer Near You*. In: CEAS. (2008).
14. <http://resources.infosecinstitute.com/cyber-kill-chain-is-a-great-idea-but-is-it-something-your-company-can-implement/>.

15. Kumar, N.: Password in practice: An usability survey. *Journal of Global Research in Computer Science*, vol. 2(5), pp.107–112. (2011).
16. Medlin, B. D., & Cazier, J. A.: An empirical investigation: Health care employee passwords and their crack times in relationship to hipaa security standards. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, vol. 2(3), pp. 39–48. (2007).
17. Meadows, D. *Thinking in Systems: A Primer*, Chelsea Green Publishing, (2008).
18. Parrish Jr, J. L., Bailey, J. L., & Courtney, J. F. *A Personality Based Model for Determining Susceptibility to Phishing Attacks*. Little Rock: University of Arkansas. (2009).
19. Cappelli, D. M., Moore, A. P., & Trzeciak, R. F.: *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes*, Addison-Wesley. (2012).
20. Kuo, C., Romanosky, S., & Cranor, L. F.: Human selection of mnemonic phrase-based passwords. In: *2nd symposium on Usable privacy and security*, pp. 67–78. ACM. (2006).
21. Sterman, J. D.: *Business dynamics: systems thinking and modeling for a complex world*, vol. 19. Boston: Irwin/McGraw-Hill. (2000).
22. Algarni, A., Xu, Y., Chan, T., & Tian, Y. C.: Toward understanding social engineering. In: *8th International Conference on Legal, Security and Privacy Issues in IT Law, (Critical Analysis and Legal Reasoning*, pp. 279–300, The International Association of IT Lawyers (IAITL). (2013).
23. Medlin, B. D., Cazier, J. A., & Foulk, D. P.: Analyzing the vulnerability of US hospitals to social engineering attacks: how many of your employees would share their password? *International Journal of Information Security and Privacy (IJISP)*, 2(3), pp. 71–83. (2008).
24. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J.: Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In: *SIGCHI Conference on Human Factors in Computing Systems*, pp. 373–382. ACM. (2010).
25. Workman, M. *Wisecrackers.: A theory-grounded investigation of phishing and pretext social engineering threats to information security*. *Journal of the American Society for Information Science and Technology*, vol. 59(4), pp. 662–674. (2008).

Internet of Things (IoT): In a Way of Smart World

Malay Bhayani, Mehul Patel and Chintan Bhatt

Abstract Internet of things-“IoT” is an interconnection of exclusively identifiable embedded computing devices where all devices are made equipped with communication and data capture capabilities so that they can use the ubiquitous internet to transmit or exchange data and other controlling purposes. IoT is expected to bring a huge leap in the field of global interconnectivity of networks. Here we are going to draw an attention on the topics which have attracted the researchers and industrialists such as remote excavation, remote mining, etc.

Keywords Internet of things · Radio frequency identification (RFID) · Long-range wireless IoT protocol (LoRa) · Lezi · Zigbee · WINEPI

1 Introduction

The popularity of IoT has been increasing greatly in the recent years due to much higher affordability and simplicity through smart devices [1]. IoT, a platform where variant networks and mass of sensors that function together and interoperate with common set of protocols. It has espoused the world through various applications like home automation, ZigBee, Big-data, and auto-id such as RFID.

Many technical communities are vigorously pursuing research topics that contribute to IoT. One of the upcoming applications is Smart ATM that can perform all the operation on user account by authenticating the user by its retina and voice. Some other embryo staged IoT applications are smart air conditioners, 3D traffic, smart building, and smart health support service. Internet of things is connecting

Malay Bhayani (✉) · Mehul Patel · Chintan Bhatt
Computer Engineering, CHARUSAT, Anand, Gujarat, India
e-mail: d14ce170@charusat.edu.in

Mehul Patel
e-mail: d14ce179@charusat.edu.in

Chintan Bhatt
e-mail: chintanbhatt.ce@charusat.ac.in

heterogeneous network/devices so that they can bring qualitative change in how we work and live. It is making our life more and more simple and increasing openness, privacy, security, analytics, and management.

2 History of IoT

Internet of things is evolved with convergence of more than one technologies [3]. The idea of smart device communication comes in 1980s but it became popular in 1990s. IoT made a revolution in technology of smart devices, wireless sensors, and networks. In the 1980s system was in existence but it did not have a name till the 1990s.

First initiative of the smart coke vendor machine was by Carnegie Mellon University which used Internet appliance to connect programmers to check cold drink in machine. After that actual rootlet of IoT track down at MIT in 1990s by the work of auto-center in networked RFID and sensing technologies. By that time, competitive congeal started for innovations in a path of IoT that we understand by some important evidence.

1999—Auto-id labs at MIT

2000—MEME (internet refrigerator) by LG

2002—Ambient orb by David rose (idea of the year by NY times)

2005—First report by ITU (UN)

2005—Nabaztag by Rafi haladzian and Olivier mevel

2008—IPSO alliance by 50 member companies including Cisco, Intel, Sap, Sun, Google

2008—Growth report of smart devices by Cisco-IBSG

2010—China plans to make major investment on IoT stat by Wen Jiabao

2011—IPv6 public launch for addressing things approximately 340 undecillion.

See Fig. 1.

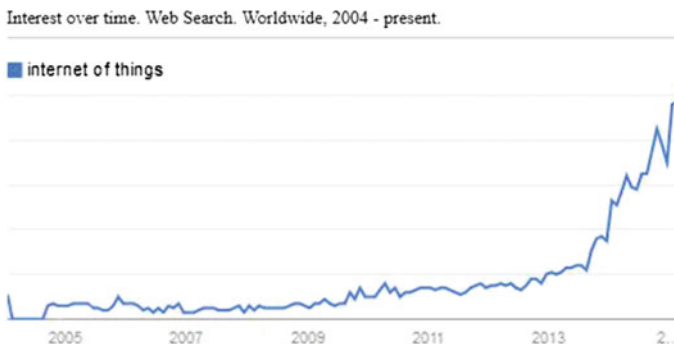


Fig. 1 IoT interest over time

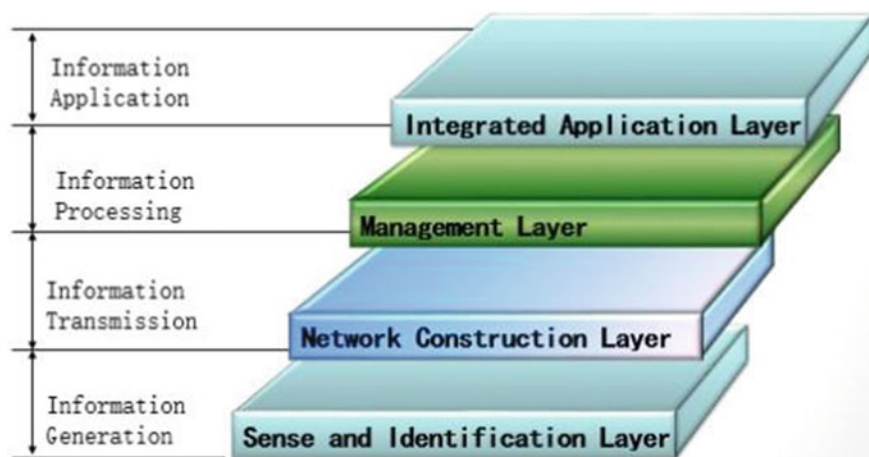


Fig. 2 IOT architecture

3 Architecture of IoT

Architecture of internet of things is designed in such a way that it can handle a large mass of data at any instance [2]. It is one of the highly reliable structure that can patch up with any element of smart device. This architecture will make us understand the abstraction of what is inherent to the actual systems. Its architecture is designed in such a way that it can be extrapolated in the reference architecture and current systems.

At its ground level, it will have sensing and identification layer. All the devices which have the ability to sense and uniquely identify that sensed data come under this layer. Generally, it is a sensor but it can also be RFID, GPS, smart devices, etc. This layer is responsible for collecting the raw data from the target, generating information, and sending it to the network construction layer (Fig. 2).

Now the layer above sensing and identification layer is network construction layer. It handles all the logistical task of the IoT. It is basically responsible for routing the information. It can be anything like WLAN, WMAN, WWAN, WPAN, Internet, etc. At the modular level WLAN is used. It is obliged to accumulate all the data from the sensing level and sending it to the management layer.

Management layer is above the network construction layer. It is designated to process all the information provided by the layer below it. All the final decisions regarding the information processing will be taken this layer. This layer may consist of the combinations of either of the given options: Data mining, information security, data center, search engine smart decision, etc. Currently, these information processing is done through cloud computing.

Application layer is the top most layer which is responsible for interpreting the data processed by the Management layer. And according to its interpretation it will

react. Suppose, if a smart home sense that temperature of your room is rising, then it will immediately turn on the air conditioner. Examples of smart applications are smart logistics, smart grid, green building, smart transport, and smart environment Monitor.

4 Algorithms of IoT

See Table 1.

5 Applications of IoT

5.1 Current Applications

As per demand and needs, existed IoT application can make seamless changes in existence of things which are related to human lives such as healthcare, transport, agriculture, energy, etc [1, 7]. When different technologies work with smart devices, sensors, networking devices at that time we recognize the value of IoT.

Smart Parking: This application is currently implemented in the city of Barcelona. Here a weight sensor is placed on each parking slot. So when the car comes and stands on it, it will get activated. When the car driver opens the mobile application, he will get to know about the number of free car parking slots on the basis of the data sent by the sensors to the cloud computer, which will in turn process the data provided by the sensor.

ZigBee: It is a description of high-level communication protocol which is used for developing PAN (personal area network). It is a wide range LAN generally termed as Smart MAN. It consumes low power and limits the transmission distance of 10–100 m line-of-sight, as per their power output and their environmental characteristic. It will be very useful to us in certain applications such as wireless light switches, traffic management, etc. Its specifications are it is inexpensive, simple to install, and initializes than other PAN like Bluetooth and Wi-Fi.

Remote Monitoring: United States is currently using this application to monitor the habitat at the Great Duck Island. They have also invested millions in installing many types of sensors in certain vegetation to track each and every movement.

LoRa (Long-range Wireless IoT Protocol): All the connected devices so far run on the same network and protocols such as Wi-Fi, Bluetooth, cellular, etc [5]. But the case of IoT device is different, they have special network requirements than smart phones, tablet, and PCs. They only need to send small data packets at a logical interval, and connect to the areas far from the traditional infrastructure of Wi-Fi and cellular. LoRa chips transmit the data in sub-gigahertz spectrum (generally it is 109, 433, 866, 915 MHz), it is unlicensed band that has less interface

Table 1 IoT algorithms

Algorithm	Description	Advantage	Disadvantage	Application
LZ78	This is a loss less data compression algorithm For each input of string it will search in its dictionary, if it is found then it work on it else it will update the dictionary	Faster decompression can be done The number of string comparisons reduces with each encoding step It attempts to work on the future data	Lz78 suffers from this problem of slow convergence There are plenty of chances that important patterns may overpass boundaries Long-term storage for the calculation of probability is not possible	Smart home Smart city Smart car and etc.
Active LeZi	This prediction algorithm helps in providing solution to the problem of information theoretic standpoint, by forecasting the upcoming symbol in sequence	It can store data for longer term for better probability It has reached the highest hit rate It has large data for predicting and its ability to pile data is high	At the time of input string parsing, all the information crossing boundaries are lost	Smart home Smart grid Smart home Appliances and etc.
Episode Discovery (WINEPI)	This algorithm manages a serial of events and also monitors its behavior and action, and helps to act on an event It identifies the frequently occurring episodes in a sequence	It helps to identify which patterns can be automated easily with least occurrence of fault	The size of episodes discovered is limited, as the window length is predefined by the user.	Smart irrigation system Smart telecommuter and etc.
Apriori	This algorithm uses breadth-first search strategy for counting item sets and generating a candidate	Easily immobilized Straightforward implementation	Difficult to find rarely appearing events Require several iterations of data Utilized fixed least support threshold	Data mining of large databases such as Banking, E-commerce
EClat	This algorithm uses vertical database layout Each item is stored together with its tidiest	Address the problem of load balancing Exploit the power of clusters or distributed systems with many nodes	In the case when it uses the existing parallel approach, it suffers from load unbalancing problem	Data mining

than others (like Wi-Fi, Bluetooth, etc.). At these frequencies, signals pierce the barriers and travel long distance.

Smart Street Light: This is the most mass energy savvy application used to control Street lights. It has sensors to detect weather and daylight. It will send the data to the data processor for analysis, in turn the street light will receive the signal of on/off lights or dim/bright light.

5.2 Upcoming Applications

Some applications are attracting the market and upcoming interesting applications have potential to lead in future with different sectors like smart ATM, remote excavation, remote mining, land slide and avalanche prevention, and chemical leakage detection in river.

6 Advantages and Disadvantages of IoT

See Table 2.

Table 2 IoT Advantages–Disadvantages

Advantages	Disadvantages
Information: To make better decision, we need to have more information. So as we all know that knowledge will help us take better and faster decisions. Suppose vegetables in the vegetable basket are going to get empty soon, so our smart basket will send us an SMS to inform us to get vegetables from the stores Tracking: Another disadvantage of IoT is tracking. It provides advance level information that could not have been possible before this so easily. Let us take an example of medical store, the application will inform the store keeper about the upcoming expiration dates of the medicines, so that they can get replaced or whatever Time: IoT saves more time which we generally used to get it wasted on gathering and processing information so that they can be accurately analyzed, in order to get better decision Money: If the cost of tagging and monitoring equipment goes down than the market for IoT will cross-skies in a very short period	Compatibility: In current time, there is no universal standard of compatibility and facility for the tagging and monitoring devices or equipment. So the disadvantages of it is that as the number and nature of devices available in market, soon it will be getting tough to connect them using IPv4 Complexity: With the help of all complex systems, there are more and more chances of failure, suppose in the vegetable market app, if the application send message about vegetable basket getting empty to two or more people with whom it is associated with them. Then both the people will go to the shop to get the vegetable as asked by app. In such a case, it may be possible that the unnecessarily double purchase of the item may be done by the people Safety: It is necessary to provide safety, else if it expired product is medicated to the patient then the ill reaction will be responded by the body and damaging health Bandwidth: It can be a problem for IoT applications, as it is limited

7 Challenges and Solutions for IoT

7.1 Challenges for IoT

According to the CISCO, there will be around 50 billion smart devices connected to the internet [2]. This figure shows that at that period of time each person the earth will be having five smart devices on an average as the prices of the processor will fall; hence it will be feasible to use processor on almost everything to make it smarter. So when these smart devices start creating data, organizations will have no organized plan to manage large data. Therefore, we need to think about where all the data generated by the processor going to be stored?

And this becomes a very serious problem. IoT promises the organizations which are going to get the insight of the customer activity. The organizations also have to maintain the data till analyzing. According to the paper published from the Gartner the Impact of the Internet of Things on Data Centers, there are several issues which have to be solved before the organization begins to earn from IoT.

The issues to be solved by organization before setting up business of IoT are:

IoT while using will generate large-scale amount of data to be processed and examined in real time, and processing large amount of IoT data will increase the workload on the data centers, thereby directing the providers to the new security, analytics, and challenges [6].

The problem is within the characteristics of IoT itself. It will connect two devices and systems and provide a data stream between the devices and the dispersed management systems. Enterprise's IT Department have to deal with IoT data as an exclusively dataset in its own. For instance, the initial set of what will build IoT data are arriving in the storage layer, same way as other unstructured data does. So, ultimately, the traditional storage architecture and management software will treat the IoT data in the same way as unstructured data.

As the number of smart devices is increasing, it will force the enterprises to bring the solution to make their system more scalable and cost-effective. Now the enterprises will have to tackle some more issues after handling the above given issues: Big number of devices, joined with sheer velocity, creates challenge, especially in the areas of security, data storage management, security, data center network, as data processing at stack.

7.2 Solutions for IoT

As far as we are concerned with the traditional storage of the data, it can be done using Hadoop. Now as we are dealing with the decreasing inefficiency of cloud computing, there is increase of burden on the cloud servers due to IoT data being processed over there. So the solution to the big data problem is to replace cloud computing with fog computing, in which all the processing and analytics works are

done on its respective routers instead of cloud servers, as a result all the data in the cloud become structured data. And the duty of the cloud server will get limited to making the data reachable the application device. For the challenge of security and privacy, we will have to increase the number of bytes being encrypted.

8 Conclusion

IoT have enormous impact on all sectors all over the world. It has helped us to improve our personal and professional life. From waking up in the morning with the help of hot coffee till switching off lights before going to sleep, we will be accompanied by the IoT. Technocrats and researcher realize revenue potential of IoT which make influence on affordable solution of problems and leads us to bright future.

References

1. Jayavardhana Gubbi, a Rajkumar Buyya, b Slaven Marusic, a Marimuthu Palaniswamia, Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions, <http://www.cloudbus.org/papers/Internet-of-Things-Vision-Future2012.pdf>.
2. John A. Stankovic, Life Fellow, IEEE, Research Directions for the Internet of Things <http://www.cs.virginia.edu/~stankovic/psfiles/IOT.pdf>.
3. Jim Chase, Texas instruments, The Evolution of Internet of Things, <http://www.ti.com/lit/ml/swrb028/swrb028.pdf>.
4. Dave Evans, The Internet of things, https://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.
5. <http://postscapec.com/home-wireless-sensor-systems>.
6. Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, Imrich Chlamtac <http://www.sciencedirect.com/science/article/pii/S1570870512000674>.
7. Understanding the Internet of Things (IoT) July 2014, http://www.gsma.com/connectedliving/wp-content/uploads/2014/08/cl_IoT_wp_07_14.pdf.

A Study of Routing Protocols for MANETs

Kalpesh A. Popat, Priyanka Sharma and Hardik Molia

Abstract A Mobile Ad hoc Network (MANET) is an autonomous set of mobile nodes, which can continue communication while moving from one place to another, without any fixed and permanent infrastructure. Every mobile node acts as a communicating node as well as forwarding node. Every node has to perform routing. Routing in wireless networks and specially in mobile networks is a tough challenge due to dynamic topologies and network partitioning problems. This paper explains routing in MANETs. This paper also surveys the unicast routing schemes to send a packet from a single source to a single destination. A conceptual comparison is given at the end of the paper to compare routing protocols.

Keywords MANET · Routing · AODV · DSR · TORA · NS2

1 Introduction

MANETs validation individual mobility and so changing topologies. As the topology is dynamic, routing is really important. Most of the routing algorithms don't cater right performance under specified scenarios where nodes are continuously dynamic their locations as easily as decent up and set [1, 2] (Fig. 1).

Figure 2 shows that whatsoever nodes in a MANET transform off due to powerfulness loser or prevent doctor by the human. In specified framework,

K.A. Popat (✉)

Marwadi Education Foundation's Group of Institutions, Rajkot, Gujarat, India
e-mail: kapopat@gmail.com

Priyanka Sharma

Raksha Shakti University, Ahmedabad, Gujarat, India
e-mail: pspriyanka@yahoo.com

Hardik Molia

Government Engineering College, Rajkot, Gujarat, India
e-mail: hardik.molia@gmail.com

Fig. 1 Node D moves out of range of A

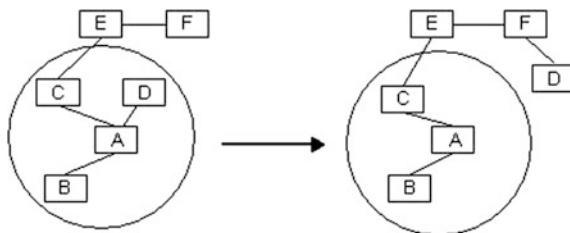
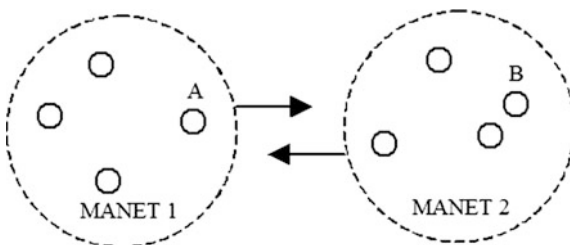


Fig. 2 MANET partitions



sometimes cloth is divided into two or many halves if the node was the only conjunctive outlet among them [2].

Plain routing keeps information most every knob in the MANET without secernent as per their locations. This strategy is fit for micro MANET to get worthy performance but it becomes ambitious as amount of nodes increases. It generates a lot of overload in maintaining message at every thickening. Hierarchical architecture divides Painter into a set of geographically unconnected miniscule chunks called the clusters. Every foregather has a set of nodes surface and one of them is elite as flock pedagogue. Routing is performed among meet heads exclusive [1, 2].

In proactive routing algorithms, so every node has perfect configuration of the network to which it belongs. Every thickening maintains stylish topology in its own database so it provides prestissimo routing. WRP and DSDV are proactive routing protocols in MANETs [1].

In unstable routing algorithms, line is searched exclusive when it is necessary. So these algorithms are pass heavy as compared to proactive algorithms but enjoin solo case when a new route is required to be created. DSR and AODV are unstable routing protocols in MANETs [1].

2 DSR—Dynamic Source Routing

Dynamic source routing protocol is a reactive protocol. Source specifies the complete concrete and full route-path to the desired destination as a part of packet header. Each intermediate node in this path works as a router and forwards the packets to the very next node given in that path. Route caching is used to cache all routers a node as has seen so far to use immediately in future. So a source first tries

to find a route from its route. If an existing route can be found, the source uses that only. Otherwise, the source tries to discover a fresh and new route by initiating route discovery process [3, 4].

As a part of the route discovery process, the source subsequently tries to flood the network with a packet asking for the route called query packet. Destination or any other intermediate nodes replies to this query packet which is stored in source's route cache. Each packet has an ID and a field to store information about a path. When a node receives a query, if it has already processed that ID or if it finds its own address in the path information, it simply discards the packet stops further broadcasting also called flooding. Else, it modifies the query message by appending its own node address in the path list and floods the query packet to the network which will to its neighbors. If a node can find route for the packet from cache, it sends a reply to the source without flooding the network then after [3].

DSR is suitable for the network in which very few numbers of nodes communicate as source nodes with very rarely used destinations. This may introduce very large end to end delay and large amount of processing overheads in very high dynamic network. Sometimes DSR is not suitable from the scalability point of view. In scalability, if the network grows, all packets like control and data become larger as they have to carry addresses of all the nodes associated in a specific path. This degrades performance because ad hoc networks are often bound by limited bandwidth [4].

3 AODV—Ad hoc On-Demand Distance Vector

The Ad hoc on-demand distance vector (AODV) is similar to DSR in a way of on-demand-based processing characteristics. It also tries to find a route on a demand basis through a similar route discovery concept. It follows a completely different mechanism to record and maintain routing data in table. It uses conventional routing table with one entry for every possible [5, 6].

A node has to flood the network with a broadcast operation of RREQ—route request message whenever it needs a fresh route. The RREQ message propagates in the network and will reach the destination for which the node is asking for the path. RREQ can also reach any of the intermediate nodes which knows the fresh path to the destination. RREP—route reply message is sent back to the original source once full fresh path is found. A node records information about all the neighboring nodes that are using each of the available routes. A link failure sends ELFN—explicit link failure notification to all the neighbors who are using that route. In DSDV, each route table entry has a destination sequenced number to resolve loop issues [5, 6].

Time-based state management is one of the most crucial tasks for AODV. It defines usability of every entry of routing table. An entry of routing table is considered expired if not used recently [6].

AODV includes an optimization for controlling the RREQ flow during route feat. It uses an expanding toroids investigate strategy to label routes for

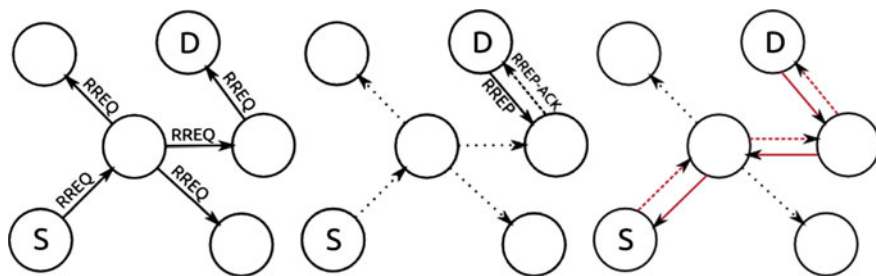


Fig. 3 RREQ and RREP processing

undiagnosed destinations. In the expanding annulus operation, monotonically increasingly larger neighborhoods are searched. The TTL theatre in the IP coping of the RREQ packets holds the depth of the seek [6] (Fig. 3).

4 TORA—Temporally Ordered Routing Algorithm

Temporally ordered routing algorithm (TORA) is an adaptive, infinite loop-free, and distributed routing scheme for wireless networks where multihop communication is needed. TORA is based on decoupling of control messages communication from the dynamic topology of the network. Link reversal algorithm is based on handling certain link failures which are structured as temporarily sequenced and based on ordered diffusing calculation. Sequence of directed link in reverse direction is stored in each of the calculations [7, 8].

Gafni and Bertsekas gave a protocol to maintain distributed destination-oriented DAG—directed acyclic graph with respect to the dynamic topology. TORA adopts this strategy. If this setup gets lost or broken due to link failures, a series of link reversals are executing to reform the destination-oriented DAG in a finite duration. TORA provides multiple paths which are guaranteed to be loop less. This is a destination-centric algorithm which has separate versions for each destination. TORA is time-consuming when link failure occurs. Once DAG is created, new links are not added and so routes may not be optimal [7].

5 DSDV—Destination Sequenced Distance Vector Algorithm

DSDV routing protocol is Bellman-Ford algorithm-based protocol. The standard DV—distance vector-based protocol, RIP—routing information protocol is based on finding shortest path among source node and destination node. RIP suffers from count-to-infinity and loop problems. In MANETs, improvised DSDV is used to

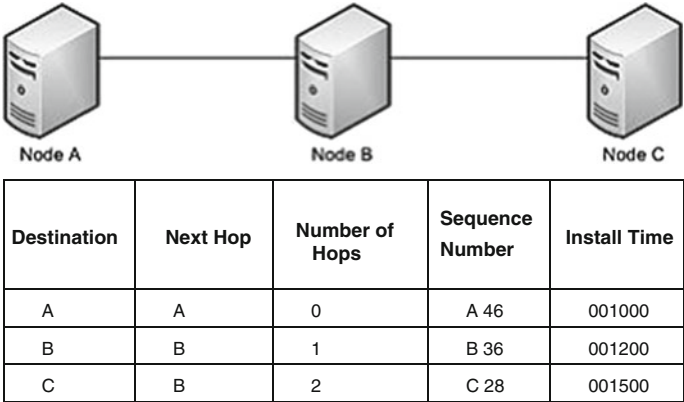


Fig. 4 DSDV table

avoid such issues. Every node’s routing table stores all destinations, the next node to go to the destination, and the total number of hopes to reach to the desired destination. DSDV propagates the changes periodically or update event based to all the neighbour nodes [9] (Fig. 4).

To avoid loop problems, every destination assigns a sequence number to the update information of another node along with the time stamp. Even numbered updates are considered as alive (present) and odd numbered updates are considered as dead (not present). Node increments its own sequence number by 2 with every advertisement and by 1 for an unreachable node (time out basis) and sends the routing information to neighbors as advertised message. Advertised information is compared by every node to node’s own routing table [9].

1. Higher destination sequence numbered route is selected.
2. If sequence numbers are equal then route with good metric is selected.

Suppose in above scenario node B finds that the node C is dead because of time out issue, it increments sequence number of C by 1 making it even. When node A advertises its own routing table to node B, B will find an entry for node C with the older sequence number. As node selects the information of higher sequence number, B continues with considering C as dead. This logic prevents count-to-infinity problem [9].

DSDV responses to the change in the topology in two ways, immediately advertisements as well as on Full or incremental update [10],

1. Immediate advertisements: Information of newly found routes, broken routes, and other metric is immediately flooded to neighbors.
2. Full or incremental update: full mode sends all information. Incremental update mode sends only modified information.

6 Simulation

Simulation is performed with NS 2.35 under MANET. Two TCP connections carrying FTP Traffic are used and analyzed independently. Various parameters are listed below (Tables 1 and 2) (Fig. 5).

Table 1 Simulation parameters

Parameter	Value	Parameter	Value
Nodes	12	Routing protocol	AODV, DSR, DSDV
Time	150 s	TCP connection 1	Node 2–8
Queue length	50	TCP connection 2	Node 5–0

Table 2 Throughput

Connection	Nodes	AODV (kbps)	DSDV (kbps)
TCP connection 1	Node 2–8	138	167
TCP connection 2	Node 5–0	142	205

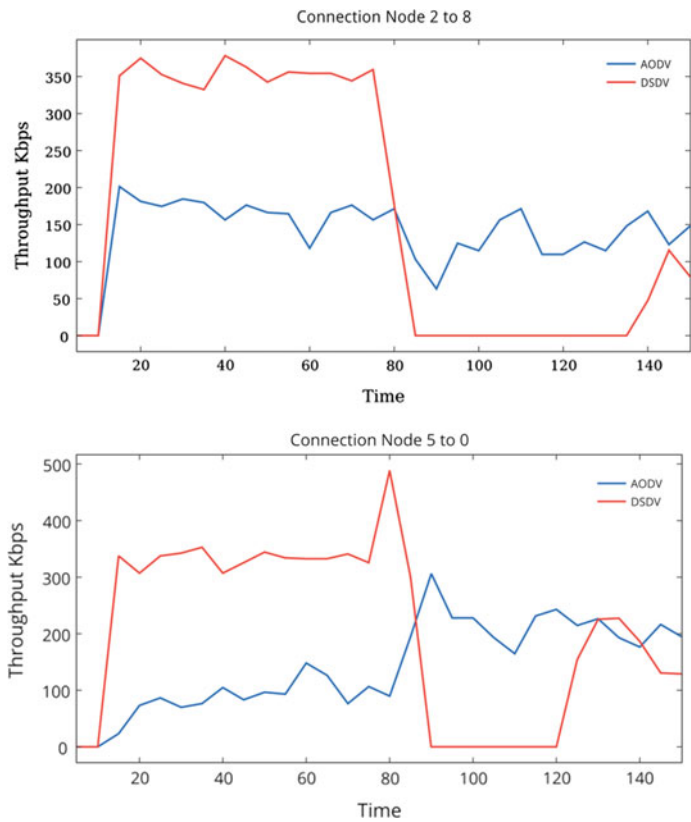


Fig. 5 Simulation results

Table 3 Comparisons

	DSR	AODV	TORA
Query	Flooding	Flooding	Flooding
Storing concept	Accumulation	Hop by hop	Hop by hop
Reply transmission	Unicast	Unicast	Flooding
Intermediate node reply	Yes	Yes	No
Multiple routers	Yes	No	Yes
Route detection	Soft	Soft	Hard

7 Conclusion

As MANETs provide mobility, routing is a difficult task. Here is a comparison of various routing protocol for MANETs. AODV and DSR are most widely used unicast routing protocols in MANETs. Effectiveness of these protocols is ultimately depends upon mobility situations. While DSR is complex from the data structure point of view, AODV is complex from the route discovery point of view. If network is highly mobile and paths are getting changed very frequently, AODV performs better. If network is comparatively less mobile and path is not getting changed for long time, DSR performs better. DSDV is the best for networks where route discovery process is not feasible. DSDV can be used if network requires immediate start of the communication without spending time in route discovery. Here is the technical comparison of reactive routing protocols (Table 3).

References

1. A.S. Tanenbaum, Computer Networks, 3rd ed., Prentice Hall, Englewood Cliffs, NJ, (1996).
2. C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, Scalable routing strategies for ad hoc wireless networks, IEEE Journal on Selected Areas in Communications, 17, 1369–1379, (1999).
3. M.S. Co rson and A. Ephremides, A distributed routing algorithm for mobile wireless networks, ACM/Baltzer Wireless Networks, 1, 61–81, (1995).
4. D.B. Johnson, Routing in Ad Hoc Networks of Mobile Hosts, Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications (WMCSA), IEEE Computer Society, Santa Cruz, CA, pp. 158–163 Dec. (1994).
5. C.E. Perkins, Ed., Ad Hoc Networking, Addison-Wesley, Reading, MA, (2000).
6. C.-K. Toh, Ad Hoc Mobile Wireless Networks, Prentice Hall PTR, Upper Saddle River, NJ, (2002).
7. R. Prakash, Unidirectional Links Prove Costly in Wireless Ad-Hoc Networks, Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications—Dial M '99, Seattle, Aug. 20, 1998, pp. 15–22.
8. F. Stajno and R.J. Anderson, The Resurrecting Duckling: Security Issues for Ad Hoc Wireless Networks, Proceedings of Security Protocols 17th International Workshop, vol. 1796 of Lecture Notes in Computer Science, Cambridge, U.K., April 19–21, 1999, pp. 172–194.

9. P. Papadimitratos and Z.J. Haas, Secure Routing for Mobile Ad Hoc Networks, SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), San Antonio, TX, Jan. 27–31, (2002).
10. L. Zhou and Z.J. Haas, Securing Ad Hoc Networks, IEEE Network Magazine, Vol. 13, no. 6, Nov./Dec. (1999).

Modelling Social Aspects of E-Agriculture in India for Semantic Web

Sasmita Pani, Priyabratta Dash and Jibitesh Mishra

Abstract There exist various web-based agriculture information systems. These systems provide the required information to farmers about different crops, soil, different farming techniques, etc. These web-based agriculture information systems deal with numerous kinds of data but they do not maintain consistency and the semantics of the data. Hence an OWL (Web Ontology Language) is used for designing required information in the web which provides meaningful annotations and vocabulary of the terms about a certain domain area to achieve the semantics for the web-based systems. Here in this paper we are building ontology of an agriculture system which is modeled in web ontology language (OWL) in protégé 5.0 framework for semantic web apps. In this paper, the usages of the farmer's or user's aspect of various components of the e-agriculture systems are analyzed, with respect to the social web components for easy access through the semantic web.

Keywords E-agriculture · Semantic web · Ontology · OWL DL · Ontology graph · Asserted model

Sasmita Pani (✉) · Priyabratta Dash
Department of Information Technology, College of Engineering and Technology,
Bhubaneswar, India
e-mail: susmitapani@gmail.com

Priyabratta Dash
e-mail: pb.dash@gmail.com

Jibitesh Mishra
Department of Computer Science and Applications, College of Engineering
and Technology, Bhubaneswar, India
e-mail: mishrajibitesh@gmail.com

1 Introduction

E-agriculture is a web-based information system that provides information to farmers/any user at any time through web. This web-based agriculture information system delivers information to users about crops, farming resources, plant nutrition, climatic conditions for a particular crop, various technologies, market information, etc. These information systems are providing resources about agricultural domain or any domain in a syntactic way. This information do not take into account the extent of usage of the user in terms of social aspects of the web components which are widely used for the social sharing of the domain information through web and these information systems do not have specified vocabulary and formal semantics of the components. Hence this web-based information system is not handling data consistency and meaningful data. To overcome these shortcomings, OWL ontologies are used in web-based applications which build semantic data for any domain. In this paper, we have made a study and analyzed the social aspects of the e-agriculture system which are beneficial to the users for accessing the different components of it, and its usage from the user's perspective.

The rest of the paper is organized as follows. Section 2 gives the overall work about e-agriculture system and ontology. Section 3 describes various phases of agriculture system, deriving ontology for an e-agriculture and designing OWL ontology for e-agriculture system. Section 4 provides an analysis among usage of user's perspectives vis-à-vis social web components through ontology graph and implementation through specifying OWL asserted model. Section 5 discusses conclusion.

2 Related Work

2.1 E-Agriculture Information Systems

The rapid advances in the technologies of wireless communications have brought opportunities for various web applications running on handheld devices. Government has started the e-Choupal project which deals with establishing internet centers in rural areas where farmers access real-time information easily. The e-Choupal portal [1] provides the rural agricultural communities with information in their respective local languages on weather forecasting, education on improved farm practices, risk management, knowledge, and purchases of better quality farm inputs. m-Krishi [2] is a high-end technical service started by Tata Consultancy Services (TCS) in 2007 in India to deliver customized advisory services to farmers on crop production, market information, and weather forecasting. m-Krishi also involves installation of different kinds of sensors in farmers field to collect information on soil humidity and weather conditions.

Mobile operator Bharti Airtel partnered with IFFCO (Indian Farmer Fertilizer Cooperative Ltd) forming the joint venture IKSL [3] in 2007. This company provides information on market prices, farming techniques, weather forecasts, rural health initiatives, fertilizer availability, etc. IKSL sends five free daily voice updates except Sunday in local language so that also illiterate farmers can be benefited. e-Sagu [4] is a tele-agriculture project started in 2004 by the International Institute of Information Technology IIIT, Hyderabad, and Media Lab Asia. e-Sagu provides farm-specific, queryless advice once a week from sowing to harvesting. This service reduces the cost of cultivation and increases farm productivity as well as the quality of agricultural products. The TNAU Agritech Portal [5], a farm technology portal has been launched in 2009 by integrating allied sectors including agriculture, horticulture, sericulture, seed sector, marketing, fisheries, forestry, and animal husbandry. The portal have feature of dynamic and multimedia-based content cover for the benefit of field extension officials and farmers in bilingual mode. It holds the information about various production technologies of agriculture crops, plant nutrition, resource management, and watershed management. This portal also holds information about agricultural engineering, agricultural marketing, and seed production.

Here Table 1 shows comparative analysis about the existing agriculture information systems.

Table 1 Comparative analysis between the various e-agriculture information systems

Information provided by these applications	e-Choupal	m-Krishi	IKSL	e-Sagu	TNAU agri PORTAL
About high yield crop	Not provided	Provided by SMS and voice specific functions in local languages through mobile apps	Not provided	Not provided	Not provided
Weather related information	Provided through computer system by ITC	Provided by SMS and voice specific functions in local languages through mobile apps	Provided by SMS and voice specific functions in local languages through mobile apps	Not provided	Provided from the web service through any handheld device
Soil	Not provided	Provided by SMS and voice specific functions in local languages through mobile apps	Not provided	Provided through experts advice in a printout form	Provided from the web service through any handheld device
Zone	Not provided	Not provided	Not provided	Not provided	Not provided

(continued)

Table 1 (continued)

Information provided by these applications	e-Choupal	m-Krishi	IKSL	e-Sagu	TNAU agri PORTAL
Seed and crop varieties	Not provided	Provided by SMS and voice specific functions in local languages through mobile apps	Not provided	Provided through experts advice in a printout form	Provided from the web service through any handheld device
Current market prices	Provided through computer system by ITC	Provided by SMS and voice specific functions in local languages through mobile apps	Provided by SMS and voice specific functions in local languages through mobile apps	Not provided	Provided from the web service through any handheld device
Fertilizers and pesticides	Not provided	Provided by SMS and voice specific functions in local languages through mobile apps	Provided by SMS and voice specific functions in local languages through mobile apps	Provided through experts advice in a printout form	Provided from the web service through any handheld device
Disease control methods	Not provided	Not provided	Not provided	Not provided	Provided from the web service through any handheld device
About interested buyers	Not provided	Not provided	Not provided	Not provided	Provided from the web service through any handheld device
Current selling market prices	Provided through computer system by ITC	Provided by SMS and voice specific functions in local languages through mobile apps	Provided by SMS and voice specific functions in local languages through mobile apps	Not provided	Provided from the web service through any handheld device
New machanisms	Not provided	Not provided	Not provided	Not provided	Provided from the web service through any handheld device

2.2 *Ontology in Web*

Ontology represents semantics, concepts, and relationships among the data in web. Ontology-driven applications [6], exhibit features such as expressiveness, extensibility, ease of sharing and reuse, and logic reasoning support. To achieve interoperability and knowledge in a shared schema, ontologies are used in any web-based application domain [7]. The ontology is the combination of classes, subclasses, axioms, relations, functions, and instances which designs data in a meaningful way. Hence ontology provides a well-founded mechanism for the representation and reasoning of information from the web [8]. Also ontology-based approaches have been used for enquiry-based learning activities in recent projects like the Concept map Learning System (CLS) and the Science Created by You (SCY) project [9].

3 *Ontology in E-Agriculture*

Nowadays OWL ontologies are used in any web-based information system which improves the information retrieval by designing the data consistently and semantically on web. It is because OWL ontologies allow building several classes, subclasses, relation/property and defining class axioms and property restrictions in any domain. The OWL ontologies can be used in e-agriculture information system to deliver semantic information to users. Before designing the data using ontology, it is necessary to analyze the various phases in an agriculture system and what information the farmer need in those phases. Here in this paper we have analyzed the various phases in an agriculture system and derived ontology in those phases which is shown in Sect. 3.1.

3.1 *Phases in an Agriculture System*

In an agriculture system, there are several phases which consist of crop identification, soil preparation and sowing, crop production and protection, harvesting and storage distribution [10]. These phases describe the requirements/information needed of a farmer to produce crop.

- Crop identification—In this phase a particular crop is selected on the basis of zone type which is again determined by the soil type and weather condition [11].
- Soil preparation and sowing—In this phase the soil bed is prepared and the sowing of the seed is done depending on seed variety by using farming equipment.

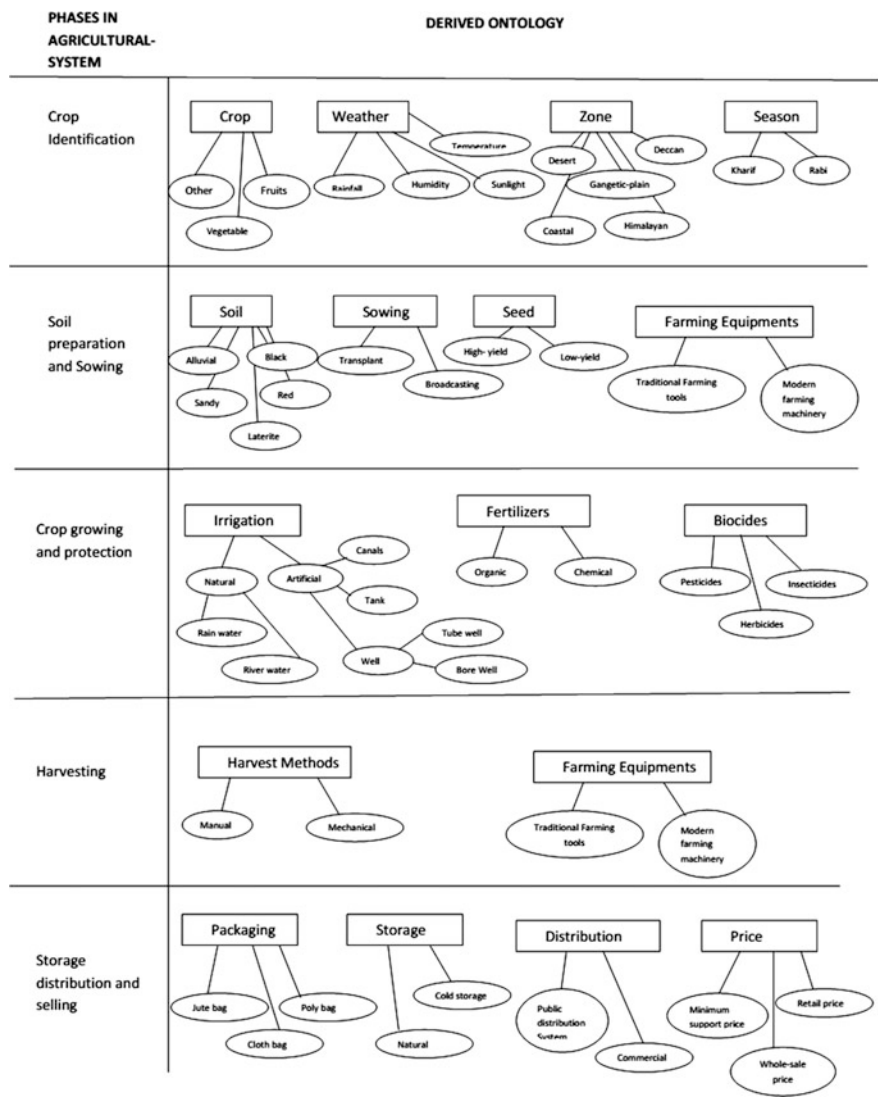


Fig. 1 Derived ontology for different phases in agricultural system

- Crop production and protection—In this phase the actual crop care and growth is monitored and is supported by activities like irrigation, applying fertilizers and biocides [12].
- Harvesting—In this phase the crop is harvested from the farm and is processed for storage and distribution.

- Storage and distribution—In this phase the harvested and processed crop is stored by packing and distributed by commercial and public distribution system [13]. The pricing is determined in the commercial distribution in terms of minimum support price (MSP), wholesale price, and retail price [14].

The derived ontology for different phases in the agricultural system is shown in Fig. 1.

4 Implementation in OWL DL

4.1 *Usage of User's View with Respect to Various Social Web Components*

The user's perspective of the various usages of the e-agriculture, with respect to the social web components are analyzed through a tabular comparative method. The farmer remains the main element which analyzes the extent of the usability of the agriculture information through the social web components like blogs, social networking, podcasting, wikis, etc. The components of user's perspective can be segregated based on the following attributes:

1. Based on role—Here the user role is analyzed based on the involvement of the farmers taking into account the agricultural components like soil, zone, crop, etc., with respect to social web components where the farmer as the user can access the information from each social web components like blogs, podcasting, content hosting, etc., which are shown in Table 2.
2. Based on preferences—Here the user preferences are analyzed based on the relevance of each of the social web components on the basis of the effectiveness, efficiency, satisfaction and memorability. The effectiveness shows the ability of a user to access the information in a particular context. Here the farmer finds the content hosting and wikis more effective on the basis of particular context than blogs, social networking and podcasting as shown in the Table 2 and Fig. 2. The efficiency provides the ability of a user to access the information in a particular context with speed and accuracy. Satisfaction is the perceived level of usage of social web components and relevance of information afforded to the user through the use of the application. Memorability shows the ability of a user to retain the use of an application effectively.
3. Based on usefulness—The users' advantage is through the usefulness of the data vis-à-vis the social web components are analyzed and based on it a particular application preferred as far as the usefulness is considered. The usefulness is determined on the basis of parameters like perceived ease of use, perceived usefulness, intention to use and trust. The farmer perceives the ease of use of blogs, social networking and wikis as more advantageous than content hosting and podcasting as shown in the Table 2.

Table 2 An analysis of user’s perspectives vis-a-vis social web components

	User’s perspectives versus social web components				
Use’s usage perspectives	Social web components				
	Blogs	Content hosting	Social networking	Podcasting	Wikis
<i>Based on role</i>					
Soil	Yes	Yes	Yes	Yes	Yes
Crop	Yes	Yes	Yes	Yes	Yes
Weather	Yes	Yes	Yes	Yes	Yes
Zone	Yes	Yes	Yes	Yes	Yes
Irrigation	Yes	Yes	Yes	Yes	Yes
Fertilizers	Yes	Yes	Yes	Yes	Yes
Farming equipments	Yes	Yes	Yes	Yes	Yes
Harvest	Yes	Yes	Yes	Yes	Yes
Storage	Yes	Yes	Yes	Yes	Yes
Distribution	Yes	Yes	Yes	Yes	Yes
Selling	Yes	Yes	Yes	Yes	Yes
<i>Based on preferences</i>					
Effectiveness	No	Yes	No	No	Yes
Efficiency	No	Yes	No	Yes	Yes
Satisfaction	No	No	No	Yes	Yes
Memorability	Yes	Yes	Yes	Yes	Yes
<i>Based on usefulness</i>					
Perceived ease or use	Yes	No	Yes	No	Yes
Perceived usefulness	Yes	Yes	Yes	Yes	Yes
Intention to use	Yes	No	Yes	No	Yes
Trust	No	Yes	Yes	Yes	No

OWL DL stands for web ontology language description logic which is a sub-language of OWL and provides logics for formal description of concepts and roles. Here concepts in ontology describe a set of individuals and role defines the relationship/property holds among them. Semantically these logics are found in predicate logics and have efficient decidability to build knowledge base information system or ontology.

4.2 Building Agriculture Semantics

In the agriculture ontology we have taken the classes such as agriculture system, crop, farmer, fertilizers, pesticides, farming equipments, seed, etc., which are semantically structured on the basis of class axioms “&owl;AllDisjointClasses.” From Table 2 we can see that from the user’s perspectives the ‘Wikis’ is the most suitable social web component, for the user’s usage in the agriculture domain. This



Fig. 2 Asserted model for user’s perspective of e-agriculture from social aspect

can be modeled semantically by taking the classes User Perspective and Social Web Component and identifying their subclasses which is analyzed from Table 2. This is shown from the asserted model in Fig. 2.

4.3 Identifying the Class Relationships and Modeling Using Ontology Graph

From Table 2, we can analyze that the subclasses under the User Perspective class are compatible with the Social Web Component class from the relationships identified as finds Usefulness In Social Web Component and prefers Social Web Component. In the former relationship the subclasses of User Perspective are satisfying the Social Web Component subclass 'Wikis'. Hence this relationship is modeled using OWL code as shown below.

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#findsUsefulnessInSocialWebComponent">
  <rdf:type rdf:resource="&owl;AsymmetricProperty"/>
  <rdf:type rdf:resource="&owl;IrreflexiveProperty"/>
  <rdfs:domain
rdf:resource="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#Effectiveness"/>
  <rdfs:domain
rdf:resource="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#Efficiency"/>
  <rdfs:domain
rdf:resource="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#Memorability"/>
  <rdfs:domain
rdf:resource="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#Satisfaction"/>
  <rdfs:range
rdf:resource="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#Wikis"/>
  <rdfs:subPropertyOf
rdf:resource="http://www.semanticweb.org/susmita/ontologies/2015/7/untitled-ontology-12#hasSocialInteractionBy"/>
</owl:ObjectProperty>
```

The above code is graphically designed in the protégé framework using the ontology graph as shown in Fig. 3. In protégé framework the ontology graph shows the relationships between the users' aspect of usage with respect to social web components which shows that user's preference toward Wikis is more than other web components as shown in Fig. 3.

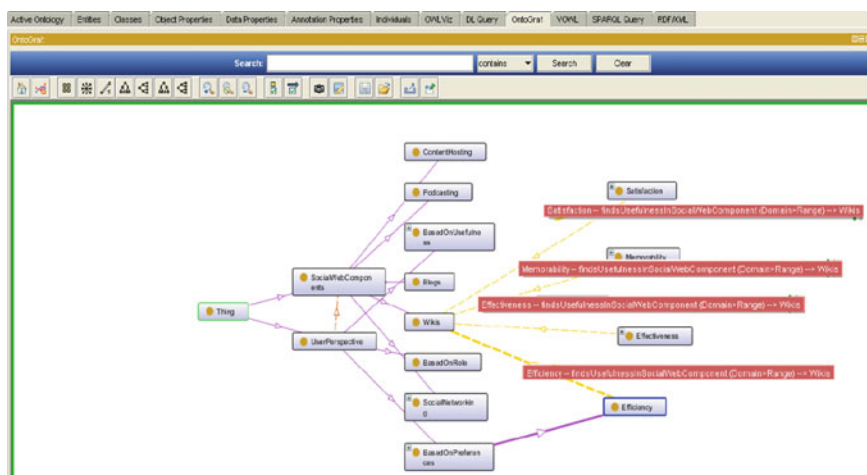


Fig. 3 Modelling the user's aspect of the social web components in ontology graph

5 Conclusion and Future Work

The previous e-agriculture systems are used to deliver information to the users as discussed and analyzed in the related works. But the data in those web-based system are not stored and arranged in a meaningful and consistent way. They do not take into user's perspective in terms of social aspects of the web components for the e-agriculture systems. The web-based information system does not handle the web-based data consistently and meaningfully.

In this paper the various classes and subclasses related to the user's perspective and the social web components of agriculture domain are identified and through the OWL language, a semantic model is created from the analysis done in Table 2. The asserted class model is verified for its consistency using Pellet reasoner in Protégé framework. It can be concluded that from the model using ontology graph in Fig. 3 that the social web component 'Wikis' has more usability from user's perspectives in the agriculture domain than other web components. In future, using this model an agricultural web portal can be developed for the semantic representation of data.

References

1. Bowonder. B., Gupta. V., Singh. A.: "Developing a rural market e-hub: the case-study of e-choupal experience of ITC", Journal in Planning commission of India, (2003).
2. Bhusan G.J.: "TCS m krishi: TCS Crop umbrella", Media lab Asia. International Journal of potato research, vol. 44, issue 2, pp. 121–125, (2010).
3. Das. V., Basu. D.: "Accessing Agricultural Information through Mobile Phone: Lessons of IKSL Services in West Bengal," Indian Research Journal of External Education, vol. 12, issue 3, pp. 102–107, (2012).

4. Glendenning J.C., Ficarelli P.P.: "The Relevance of Content in ICT Initiatives in Indian agriculture", IFPRI discussion paper, (2012).
5. TNAU agritech portal: www.agritech.tnau.ac.in.
6. Ahmed S., Parsons D.: "An Ontology-Driven Mobile Web Application for Science Enquiry Based Learning", In: 7th International conference on Information Technology and applications, pp. 255–260, (2011).
7. Zhdanova A. V., Ning L., Moessner K.: "Semantic Web in Ubiquitous Mobile Communications", *The Semantic Web for Knowledge and Data Management: Technologies and Practices*, (2009).
8. Nan X., Zhang W. S., Yang H. D., Zhang X. G., Xing X.: "CACOnt: a ontology-based model for context modeling and reasoning." *Applied Mechanics and Materials*, vol. 347, pp. 2304–2310, (2013).
9. Sohaib A., Parsons D.: "ThinknLearn: An ontology-driven mobile web application for science enquiry based learning", In: 7th International Conference on Information Technology and Application, pp. 255–260. (2011).
10. The Agricultural Promotion and Investment Corp. Of Orissa Ltd. (apicol), Bhubaneswar, orissa, India.
11. Misri B. K.: "Country Pasture/Forage Resource Profiles", India.
12. State of Indian Agriculture, Directorate of Economics and Statistics, Ministry of Agriculture, New Delhi, India, (2013).
13. Mitra S., Sareen J.S., "Adaptive policy case study: agricultural price policy in India", In: *Designing Policies in a World of Uncertainty, Change and Surprise: Adaptive Policy-making for Agriculture and Water Resources in the Face of Climate Change—Phase I Research Report*, (2006).
14. The Pocket book on agricultural statistics, Directorate of Economics and Statistics, Ministry of Agriculture, New Delhi, India, (2013).

An Efficient Adaptive Data Hiding Scheme for Image Steganography

Sumeet Kaur, Savina Bansal and R.K. Bansal

Abstract Steganography is crucial for maintaining integrity, authenticity, copyright protection, illegal use detection, and distribution of digital media over public networks. In this paper, a new adaptive steganography technique is proposed to provide better tradeoff between the two main conflicting parameters—capacity and robustness. An adaptive approach is exploited to decide about the required position and size of the cover image pixels for data hiding to achieve the desired QoS parameter. Performance comparison of the proposed scheme in terms of imperceptibility, capacity, PSNR, and RMSE w.r.t. the PVD-based steganography technique proves it to be an efficient technique.

Keywords Secret information · Security · Steganalysis · Steganography

1 Introduction

With the growth and development of the Internet, there is need for security tools to provide secure communications over public networks. One of the objectives of steganography is to hide information into digital media without any detectable trace in such a way that none other than the sender and the intended receiver know the presence of hidden secret information. Steganography is an art and science of embedding secret message into an appropriate carrier object like image, video, sound, or other file [1, 2]. Embedding may be parameterized by a key that makes it difficult to even detect the presence of data. Once cover object is embedded with a secret message, it is called the stego object.

Steganography conceals the existence of hidden secret data, while cryptography scrambles message so that it cannot be understood though the cipher text generated

Sumeet Kaur (✉)
Punjabi University, Patiala, India
e-mail: purbasumeet@yahoo.co.in

Savina Bansal · R.K. Bansal
Maharaja Ranjit Singh State Technical University, Bathinda, India

arouses suspicion [3]. The main motivation behind the use and design of steganography and watermarking techniques is to protect digital media such as books, software, music, film, etc., from unauthorized use and distribution [4].

Interest in the field of steganography is now greater than before due to increased use of Internet for communication and other data transfer purposes such as for handling business, commercial, and other financial transactions. There is a huge demand for steganography protocols due to its vast number of applications such as confidential transmission, video surveillance, military and medical applications, band captioning, integration of multiple media for convenient and reliable storage, management, transmission, embedding executables for function control, error correction, and version upgrading, etc. Mainly, publishing and broadcast industries are now interested in the use of steganography and watermarking techniques for hiding serial number and marks for copyright protection [5].

The technical challenge in steganography is to find suitable redundant bits in cover object to embed secret data and also show sufficient resistance to various attacks and transformations [6]. The primary goal of steganography techniques is to maximize embedding rate and minimize the delectability of the resulting stego images against steganalysis techniques [7].

2 File Formats and Type of Compression Used

Over the Internet BMP, JPEG, and GIF are common file formats. BMP images mostly preferred are lossless 24 bit images; the next best format is 256 color and grayscale images like GIF files for steganography. The type of compression used plays an important role in steganography. There are two types of compression techniques in use:

- Lossless compression: When we require original information to remain intact, lossless compression is preferred. This type of compression is supported by GIF and BMP file formats.
- Lossy compression: To save space, a good amount of compression is used but integrity of original file may not be certain. JPEG supports this type of compression.

3 Steganography Techniques

Steganography techniques are classified as spatial domain and transform domain techniques [8]. Some of the commonly used methods to manipulate the cover object to hide secret data are spread spectrum, masking [9], statistical, and distortion techniques.

- Spatial domain techniques: Spatial domain techniques include bit insertion and noise manipulation. These techniques are simple and provide a high level of capacity but are less robust and require lossless image formats.
- Transform domain techniques: Transform domain techniques involve various image transforms such as DCT and wavelets [10, 11]. These techniques are more robust but have less capacity. Transform domain techniques are generally preferred for watermarking purposes, where the focus is on robustness rather than capacity.

4 Steganalysis

Steganalysis is the art and science of detecting hidden messages. Research in steganalysis not only provides ways to detect hidden information but also provides motivation to improve steganography methods. Steganalysis techniques are divided into two broad categories:

- Universal or blind steganalysis: Universal techniques are a general class of algorithms that work for a range of steganography algorithms.
- Target or specific steganalysis: Specific steganalysis techniques are designed for a specific type of steganography algorithm.

These steganalysis techniques work further on two types of approaches: statistical steganalysis and feature-based steganalysis. Statistical steganalysis use spatial or transform domain methods to detect the presence of hidden message, while feature-based steganalysis uses feature of cover to detect the presence of hidden data [12–14].

5 Desired Parameters for an Efficient Steganography Algorithm

Embedding capacity and robustness are desired parameters for an efficient steganography algorithm, but there is tradeoff between these two parameters [1, 15].

- Statistical undetectability (Imperceptibility) of the data: It determines how difficult it is to detect the presence of hidden data. The higher the stego image quality, the more invisible the hidden message.
- Steganographic capacity: It is the maximum length of secret data that we can embed in cover object without affecting any visual detectability and statistical properties of the given object.

- **Robustness:** It refers to how well the steganographic system will be able to resist steganalysis attacks to prevent extraction and modification of hidden secret data. Watermarks are an example of a robust steganography technique.

6 Limitation in Existing Techniques and Motivation for the Work

A large number of steganography techniques have been proposed by researchers from time to time. However, most of the existing schemes get attacked by steganalysis methods. LSB-based techniques that exploit least significant bits under the assumption that LSB planes are insignificant and random do not hold good especially for images with smooth regions. Further, most of the embedding techniques [10, 11, 16–21] suffer from the following limitations:

- Change in visual quality of cover object
- Change of statistical properties of cover object
- Introduction of noise and detectable fingerprints
- Change in the size of original file, increase in number of colors, intensity, etc.

These indications and patterns can be attacked by steganalysis. All these problems need to be addressed while designing an efficient and robust steganography system.

7 Proposed Scheme

To overcome the various limitations, we propose an efficient embedding scheme that uses gray image as the underlying cover object. The major steps involved in designing an embedding algorithm are as follows:

- (1) The text file to be hidden is created and converted into equivalent ASCII form. Finally text file in binary form is generated. Binary file is scrambled by segmenting it into different parts and changing their order to form a new binary file to provide additional layer of security.
- (2) An adaptable range table is then prepared based on the different intensity values for pixels of cover image. Here cover image is considered as an 8-bit grayscale image and pixels having intensities in the range 0–255. To improve imperceptibility, the number of bits selected for each pixel in each range are varied and more bits are selected for pixels having higher intensity (edge areas) and fewer bits are selected for pixels having lower intensity (smooth areas). Range table is prepared by the following rules:

- (a) For pixels having intensity in range 0–192, one bit is selected for each pixel to embed secret data.
 - (b) For pixels having intensity of range 193–223, two bits are selected for each pixel to embed secret data.
 - (c) For pixels having intensity of range 224–239, three bits are selected for each pixel to embed secret data.
 - (d) For pixels having intensity of range 240–255, four bits are selected for each pixel to embed secret data.
- (3) Pixels are selected as per the QoS required, and can be adapted depending on the requirement of applications as capacity and robustness.
 - (4) To increase the security of the hidden text a zigzag scan pattern is used to select the pixel of cover image to defeat steganalysis process.
 - (5) The above steps are repeated until the whole of the secret data is embedded.

Further extracting procedure at receiver side involve reverse steps to get secret data from the stego image.

8 Result and Analysis

To check the embedding capacity of the proposed steganography algorithm various experiments are performed with different type and size of hidden secret message. Further visual and statistical analysis is performed on benchmark images as Lena, Baboon, and Cameraman to check the efficiency of proposed algorithm.

Visual analysis: After hiding different size and types of data in cover images visual analysis is made to find out any distortion with stego images.

As shown in Fig. 1, cover images and their respective stego images are visually analyzed and it is observed that there is no dissimilarity in cover and stego images and no distortion is detected visually after embedding.

- *Statistical analysis:* Statistical analysis is performed to check changes in cover images after embedding. For statistical analysis, various parameters considered are PSNR (Peak Signal to Noise Ratio) and RMSE (Root Mean Sq. Error) are calculated for different cover and stego images. Results are represented in tabular form in Table 1.

The proposed technique is compared with well-known existing methods and comparison results are shown in Table 2.

Results in Table 2 show that the proposed scheme not only has good capacity but also high imperceptibility.

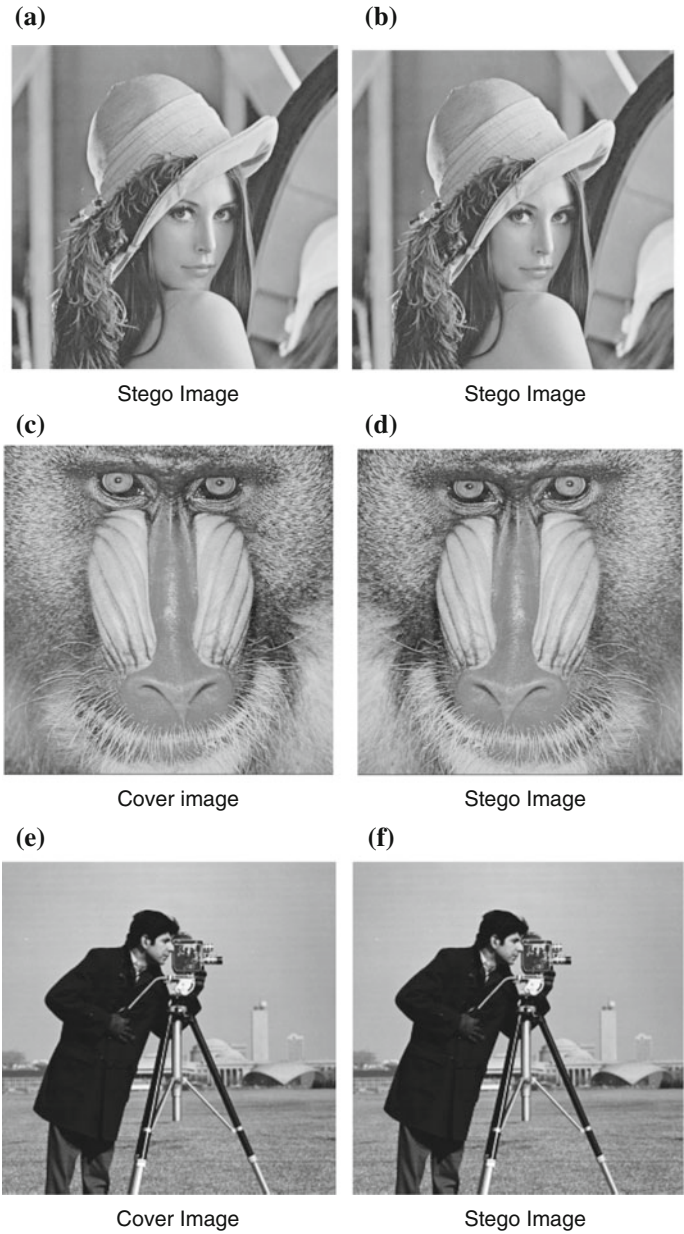


Fig. 1 **a** Cover image of Lena. **b** Stego image of Lena. **c** Cover image of Baboon. **d** Stego image of Baboon. **e** Cover image of Cameraman. **f** Stego image of Cameraman

Table 1 Capacity, PSNR, and RMSE for proposed scheme for different benchmark images

Image (size 512 * 512 bytes)	Size of hidden secret data (in bits)	PSNR	RMSE
Lena (512 × 512)	87,990	63.8640	0.0267
Baboon (512 × 512)	87,990	64.0464	0.0256
Cameraman (512 × 512)	87,990	64.3022	0.0241

Table 2 Comparisons of proposed method with other methods

Cover images (512 * 512)	PVD method [22]	Tri-way pixel method [23]	Proposed method
	<i>Capacity (bytes)</i> <i>PSNR (dB)</i>	<i>Capacity (bytes)</i> <i>PSNR (dB)</i>	<i>Capacity (bytes)</i> <i>PSNR (dB)</i>
Lena	50,960 41.79	75,836 38.89	87,990 63.8640
Baboon	56,291 37.90	82,407 33.93	87,990 64.0464
Peppers	50,685 40.97	75,579 38.50	87,990 64.3022

9 Features of Proposed Technique

Although the main objective of the proposed technique is to provide high embedding capacity, at the same time certain attempts are also made to design a robust technique simultaneously. It is tried to achieve optimum balance between conflicting but desirable parameters as capacity and robustness. The proposed scheme has high embedding capacity, very high value for peak to signal ratio, and proves validity and efficiency of scheme with low root mean square error. Further visual analysis gives no distortion of stego images and is capable of preserving statistical properties and defeating steganalysis attacks.

10 Conclusion

Steganography is a growing field with a vast number of applications. From the past decade various techniques have been designed with their own pros and cons. Yet, there is need to design high capacity techniques capable of resisting steganalysis attacks. Research in this field is also motivated and driven by forces to eliminate limitations of other security-related technologies and with it own objective to provide secure communication over the Internet. Performance analysis of the proposed steganography scheme in terms of imperceptibility, capacity, PSNR, and RMSE parameters proves it to be an efficient technique.

References

1. Bender, W., Gruhl, D., Morimoto, N. & Lu, A.: Techniques for data hiding, IBM Systems Journal, Vol 35, No. 3–4, pp 313–316 (1996).
2. Lee, Y.K. and Chen, L.H.: High capacity image steganographic model, Visual Image Signal Processing, 147:03, (2000).
3. Stefan Katzenbeisser, Fabien A. P. Petitcolas: Information hiding techniques for steganography and digital watermarking, Artech House Books, ISBN 1-58053-035-4 (December 1999).
4. Fabin A.P. Petitcolas, Ross J. Anderson and Markus G. Kuhn: Information Hiding—A Survey, Proceedings of the IEEE, special issue on protection of multimedia contents, pp 1062–1078 (1999).
5. Fabin A.P. Petitcolas, Ross J. Anderson: On the Limits of Steganography, IEEE Journal of Selected Areas in Communications, 16(4):474–481, Special Issue on Copyright & Privacy Protection. ISSN 0733-8716 (1998).
6. Weiqi Luo, Fangjun Huang, and Jiwu Huang: A more secure steganography based on adaptive pixel value differencing scheme, Multimedia Tools and Applications, Springer, (Jan 2010).
7. M. Kharrazi, H.T. Sencar and N. Memon: Cover Selection for Steganographic Embedding, IEEE International Conference on Image processing, Atlanta USA, pp 117–120 (2006).
8. Sumeet Kaur, Savina Bansal, R.K. Bansal: Steganography: Concepts and Practice, 7th International Conference on Advanced Computing and Communication Technologies at APIIT SD India, Panipat, in association with IETE, Delhi and technically cosponsored by Computer Society IEEE Delhi, pp 113–119 (2013).
9. Marvel, L.M., Bonchelet Jr., C.G. & Retter, C.: Spread Spectrum Steganography, IEEE transactions on image processing, 8:08 (1999).
10. P. Amat, W. Puech, S. Druon, J.P. Pedebay: Lossless 3D Steganography based on MST and Connectivity modification Signal Processing: Image communication Volume 25, Issue 6 Elsevier, pp 400–412 (2010).
11. Z. Ni, Y. Q. Shi, N. Ansari, and W. Su: Reversible Data Hiding, IEEE Transactions on Circuits and Systems for Video Technology, VOL. 16, NO. 3, pp 354–362 (2006).
12. Sherif M. Badr and et al: A Review on Steganalysis Techniques: From Image Format Point of View, International Journal of Computer Applications (0975—8887), Volume 102—No. 4, pp 11–19 (September 2014).
13. N.F. Johnson and S. Jajodia: Steganalysis of Images Created Using Current Steganographic Software, Proceedings of 2nd Int'l Workshop in Information Hiding, Springer-Verlag, pp. 273–289 (1998).
14. Westfield and A. Pfitzmann: Attacks on Steganographic Systems, Proceedings of Information Hiding—3rd Int'l Workshop, Springer Verlag, pp. 61–76 (1999).
15. Provos N. & Honeyman P., Hide and Seek: An introduction to steganography, IEEE Security and Privacy Journal, pp 32–44 (2003).
16. Weiqi Luo, Fangjun Huang, and Jiwu Huang: Edge Adaptive Image Steganography Based on LSB Matching Revisited, IEEE Transactions On Information Forensics And Security, Vol. 5, No. 2, pp 201–214 (June 2010).
17. Z. Zhao, N. Yu, and X. Li, : A novel video watermarking scheme in compression domain based on fast motion estimation, Proceedings of IEEE International Conference on Communication Technology, pp. 1878–1882 (2003).
18. Chin-Chen Chang, Hsien-Wen Teng: Steganographic method for digital images using side match, Pattern Recognition letters, volume 25, pp 1431–1437 (2004).
19. Jarmo Mielikainen: LSB matching revisited, IEEE Signal Processing Letter Vol 13, No. 5, pp 285–287 (2006).
20. Weiqi Luo, Fangjun Hung, Jiwu Hung: Edge Adaptive image Steganography based on LSB Matching Revisited, IEEE Transaction of information Security, Vol 5, No. 2, pp 201–214 (2010).

21. Wen-Jan Chen, Chin-Chen Chang, T. Hoang Ngan Le: High Payload Steganography Mechanism Using Hybrid Edge Detector, Expert System With Applications Volume 37, Issue 4, Pages 3292–3301(2010).
22. D.C. Wu and L.C Tsai: A Steganography methods for images by pixel value Differencing, Pattern Recognition Letters, Vol. 24, pp. 1613–1626 (2003).
23. Ko-Chin Chang, C-P. Chien-Ping Chang Ping S. Huang and Te-Ming Tu.: A Novel image Steganographic Method Using Tri- way Pixel value Differencing, Journal of Multimedia, Vol. 3, No. 2, pp 37–44 (June, 2008).

A Five-Layer Framework for Organizational Knowledge Management

H.R. Vishwakarma, B.K. Tripathy and D.P. Kothari

Abstract From a strategic point of view, the most valuable assets in the present era are knowledge assets. Effective management of knowledge assets is essential to use existing knowledge bases as well as to create new knowledge bases by knowledge workers. Therefore, knowledge assets and knowledge workers are crucial aspects of any Knowledge Management (KM) Framework. Most of the researchers have viewed knowledge management either from technical or management perspectives. From architectural perspectives either a centralized or a peer-to-peer approach has been mentioned in the literature. We present a holistic view of knowledge management considering its multiple dimensions and stakeholders to achieve specific organizational knowledge vision and goals. In this paper, we propose a five-layer knowledge management framework comprising of knowledge vision, knowledge processes and services, knowledge networks, knowledge asset management, and knowledge assets. Also, we describe an architecture for interaction among knowledge workers.

Keywords Knowledge management framework · Knowledge process · Knowledge networks · Knowledge assets · Knowledge sharing community · Knowledge services

H.R. Vishwakarma (✉) · B.K. Tripathy · D.P. Kothari
VIT University, Vellore 632014, Tamil Nadu, India
e-mail: hrvishwakarma@vit.ac.in

B.K. Tripathy
e-mail: tripathybk@vit.ac.in

D.P. Kothari
e-mail: dpk0710@yahoo.com

1 Introduction

There has been a widespread development/deployment of knowledge management (KM) systems. The acquisition and assimilation of knowledge is a major thrust area of organizational processes to realize vision and strategic goals of organizations. Survival in this knowledge era will largely depend on how organizations gear up to: capitalize on individual know-how in a collective knowledge; improve newcomer learning and integration; disseminate best practices; improve work processes, product quality and productivity; and reduce time frame for new product/solution design. The above requires knowledge workers, knowledge work products, knowledge networks, knowledge processors/servers, and various other bodies of knowledge.

Binney [1] presents current knowledge management theories along with their applications, associated tools and technologies described in the literature. Liao [2] classifies knowledge management technologies into the following seven categories: knowledge management framework, knowledge-based systems, data mining, information and communication technology, artificial intelligence/expert systems, database technology, and modeling. Hoegl and Schulze [3] discuss ten knowledge management methods and how these methods support knowledge creation during the development of new products.

Maier and Hädrich [4] present two knowledge management system architectures, out of which one is centralized and the other one is peer-to-peer (P2P) architecture. They also outline differences between these architectures. They consider centralized approach as inappropriate and ineffective for knowledge sharing, even though most of current knowledge management systems are based on the centralized network structure. On the contrary, a P2P architecture enables and supports collaboration people, groups of individuals, and organizations to work together to accomplish a task or a collection of tasks. Baslen et al. [5] discuss how the networks of practice emerged and how knowledge portals stimulate knowledge sharing.

According to Tiwana and Bush [6], the primary objective of forming knowledge networks is to develop, distribute, and apply knowledge. Thus, these networks help in harnessing distributed expertise. Kim [7] proposes a knowledge grid/P2P architecture that supports three types of workflow knowledge models and a configurable physical architecture. Kwok and Gao [8] discuss how a virtual knowledge sharing community can benefit from decentralized P2P technology. Haase et al. [9] propose a model that helps in exploring semantic similarity between the subject of a query posed by an individual and the advertised expertise of other peers. Thus, a peer can select appropriate peers to get a query answered.

Nerkar and Paruchuri [10] suggest the likelihood of knowledge usage is determined by the characteristics of knowledge inventing positions in an intra-organizational network of inventors or intra-firm knowledge network. A study by Boh [11] shows that there are two key factors which help users to overcome difficulties in reusing knowledge assets: seeking assistance from and sharing a common perspective with the author of the knowledge asset. Li and Chang [12]

propose a solution integrating text extractor, slideshow generator, knowledge repository, etc., to share presentational knowledge assets.

Despite several search engines and indexing techniques, many knowledge seekers prefer asking their queries from a human expert rather than by searching online sources. There are many advantages of knowledge sharing through direct interactions between a knowledge seeker and an expert rather than sharing knowledge by codifying it from any expert. Sometimes, an individual may like to use the knowledge available in document form, through conversations, or in meetings.

A knowledge network usually contributes to the effectiveness and efficiency of an individual in teams. Also, the knowledge sourcing behavior determines how an individual gains knowledge from others. Further, a team learning and productivity is influenced by the factors like team stability, team member familiarity, and interpersonal trust. New tools or channels of communication may also be needed to facilitate knowledge sharing among knowledge workers, for example, a tool to answer questions about the organization's knowledge repositories, knowledge processes or services, knowledge networks or communities of practitioners.

In this paper, we propose a model for knowledge management that begins with the articulation of organizational knowledge vision and strategic goals. We feel that organizational knowledge management should be viewed from three dimensions, viz., people, process, and technology. We present a centralized approach for knowledge process/service management and knowledge asset management, whereas distributed approaches for knowledge creation, acquisition, sharing, and evaluation.

The outline of remainder part of the paper is as follows. Section 2 discusses a five-layer framework organizational knowledge management. Section 3 presents the system architecture of user–expert interaction. Section 4 lists the advantages of the system.

2 A Five-Layer Knowledge Management Framework

We propose a five-layer framework for organizational knowledge management as depicted in Fig. 1. It comprises of knowledge vision, knowledge processes and services, knowledge networks, knowledge assets management, and knowledge assets accessible to knowledge workers. The functionalities of various layers along with the actors and their roles are given in Table 1. This framework is suitable for both intra- and inter-organizational scenarios. Further, individual knowledge workers or organizations could benefit from collaborative knowledge management.

Knowledge vision is articulated by the top management. The mission, strategic goals, and domain of an organization are determined by knowledge vision. An organization gets competitive advantage based on knowledge and skills of its employees. The above enables an organization to create new products/processes/services, or improve existing ones more efficiently and/or effectively.



Fig. 1 A five-layer framework for organizational knowledge management

Table 1 Framework layers along with functionalities, actors and their roles

Layers	Functionalities	Actors and their roles
Knowledge vision	Provision of articulation/update of organizational knowledge vision and strategic goals	Super-knowledge worker articulates vision and strategic goals as well as communicates these to other stakeholders
Knowledge processes and services	Process management, process improvement, services portfolio management, service quality and level management	Super-knowledge workers, experts and supervisors define process and services as well as formulate/enforce knowledge management policies and guidelines
Knowledge networks	Management functionalities of knowledge networks, and access privileges	Expert knowledge workers along with supervisors form and manage knowledge networks, invite knowledge workers to join knowledge networks
Knowledge asset management	Management functionalities of knowledge assets and access privileges	Knowledge asset owners/creators, and supervisors manage knowledge assets and suggest guidelines for publishing/sharing knowledge asset management policies
Knowledge assets	Repository of knowledge assets	Individual knowledge workers create and/or own knowledge assets

Knowledge processes and services represent activities such as processing/compiling knowledge extracted from different sources, one or more knowledge bases, and support for storage and retrieval of knowledge instances. These might support and/or include query processors and search engines—including desktop, intranet, and the Internet search engines.

Knowledge workers are workers whose main capital is knowledge. They are persons employed to produce or analyze ideas and information. They can assume the role of knowledge producer or consumer or both. A few super-knowledge workers may assume the role of active managers for organizational knowledge. Sometimes, they play the role of neutral facilitators of knowledge sharing among individual peers.

The importance of coordination, learning/innovation, translation/local adaptation, and support for individual knowledge workers cannot be overemphasized. Knowledge networks are formed to cater to the above need. These networks play two important roles. One of which is about supporting both personal and organizational knowledge management. The other one is about enabling individual knowledge workers to share knowledge, and helping organizational memory to grow.

Knowledge assets management is crucial for the very survival of organizations given the nature and characteristics of knowledge assets. Knowledge workers and the knowledge networks formed by them play crucial roles in managing knowledge assets. Knowledge workers create and/or own knowledge asset as well as use knowledge assets.

In this paper, we consider knowledge networks as collections of interdependent individuals and teams who come together across organizational, spatial, and disciplinary boundaries to create and share knowledge. They typically share a body of knowledge and common communication channels. The primary resources used, shared, and outcome produced by them are knowledge assets. The following section illustrates some of these aspects.

3 System Architecture for User–Expert Interaction

As aforesaid, knowledge workers assume various roles in an organization and interact among themselves to carry out their individual tasks and/or team-based tasks. A knowledge worker might play the role of an expert in one scenario or a user in another scenario. Knowledge users and experts constitute a knowledge network depending on knowledge sharing needs and frequency of interactions. In this section, we propose architecture of a system to enable such interactions. The proposed system dynamically manages pairs of questions and answers. A question is either answered by the system or forwarded to an expert. A session of interactive questions and answers may be facilitated between a user and an expert. A question can be broken into sub-questions thus creating a set of probing questions related to a topic chosen by a user. An answer provided by the expert is presented to the user as well as filed for future enquiries. An answer might to be associated with a document containing rationale and explanation.

Figure 2 depicts the system architecture comprising of five main functional modules, viz., User Module, Expert Module, Supervisor Module, Filer/Retriever, and User–Expert Interaction Module.

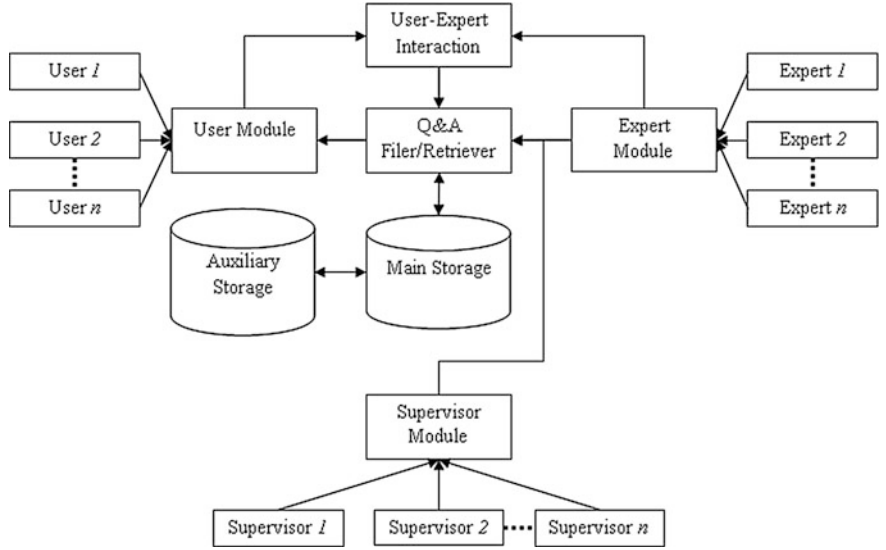


Fig. 2 System architecture for user-expert interaction

An end user is essentially the knowledge seeker/user. An expert, is the one who is knowledgeable and experienced in the particular subject, Expert users are responsible to update the InfoBase/Knowledge base. Supervisor or Super-knowledge worker is responsible for update of the expert profile, end user profile, and authorization database.

Figure 3a–c illustrate three modules one each for users, experts, and supervisors respectively. Users can browse the portions of InfoBase and NoticesDB matching to their profiles and privileges. Also the system broadcasts various messages to the appropriate users or groups of users as and when triggered by the experts or the supervisor.

The enquiry handler scans the InfoBase as per user question and displays answer in case of successful match; else it forwards the question to an appropriate expert. Also, the enquiry handler facilitates a direct user-expert interaction on request.

Apart from answering questions, experts can modify pairs of questions and answers based upon users’ feedback. The Expert module allows changes to InfoBase/Knowledge base, NoticesDB, and Broadcast Messages Database initiated by the expert concerned with a particular subject. The Supervisor module facilitates changes to Expert profile, User profile, and Authorization Databases initiated by the supervisor.

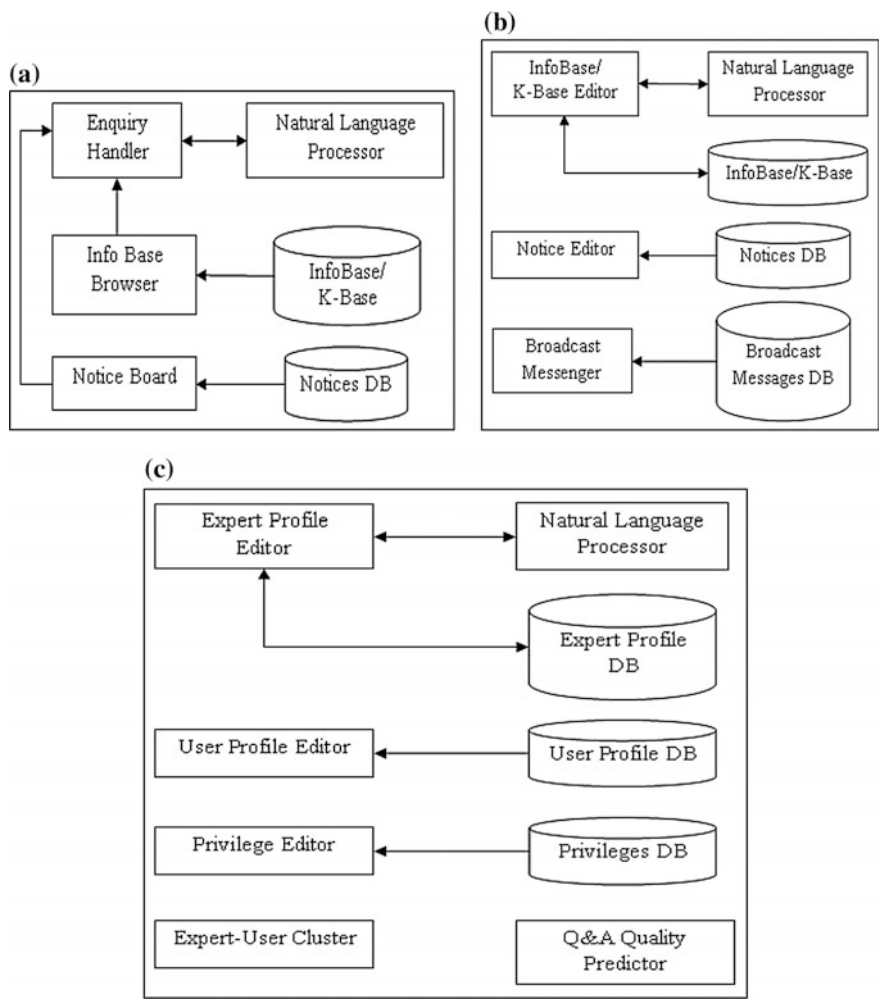


Fig. 3 a User module. b Expert module. c Supervisor module

4 The Advantages of the Proposed System

We present here the advantages offered by the proposed system. These can be grouped in three categories one each pertaining to users, experts, and organizations.

(a) The distinct benefits for the end users:

- Authentic and faster answers to questions and associated details.
- No restrictions on type and number of questions.
- No restriction of time, answers are provided even in the absence of an expert.

- Even scattered and diverse information is made available at a single point.
 - Resistance to share information is minimized.
 - They get the right piece of information themselves.
 - No psychological barrier of shyness in asking questions.
 - Better inquisitiveness as the knowledge seekers need not disclose their identity.
 - Overall culture change for knowledge sharing.
- (b) The distinct benefits for the experts:
- No need to answer the repetitive questions.
 - More time to develop new knowledge.
 - Motivation to increase their expertise as the same is used continuously.
- (c) The distinct benefits for the organizations:
- Better productivity due to faster knowledge sharing.
 - Consistent and authentic information.
 - Better utilization of time and efforts of the experts.
 - No sudden stoppage in the flow of information, even when experts resign or retire.
 - Even tacit knowledge is shared systematically.
 - As experts get recognition and visibility, quality of information improves.
 - Support for virtual knowledge sharing communities and personal learning networks.

In a nutshell, end users get answers to their questions at a single point without barrier of time and availability of experts in person. Experts become more productive as they need not answer repetitive questions. Organizations benefit from optimal use of knowledge and human capital. The interdependency of the above helps grow the knowledge networks thus improving organizations' collective knowledge. Further, this system could be used for building personal learning networks, especially for newly joined members in a team or an organization.

5 Conclusion

This paper summarized previous literature and approaches on knowledge management. It suggested a holistic approach considering multiple dimensions and stakeholders involved. The paper proposed a five-layer framework for organizational knowledge management from multiple perspectives. It also discussed the system architecture to facilitate interaction among knowledge workers and to improve organizational knowledge. The system can (1) enhance productivity of knowledge workers (2) improve quality of organizational knowledge, and (3) ensure growth of knowledge networks. Further work can be done to study the impact of user–expert interaction on knowledge flow dynamics.

References

1. Derek Binney: The knowledge management spectrum understanding the KM Landscape. *Journal of Knowledge Management*, Volume 5, Number 1, pp 33–42 (2001).
2. Shu-hsien Liao: Knowledge management technologies and applications—literature review from 1995 to 2002. *Expert Systems with Applications* 25 (2003) 155–164.
3. Martin Hoegl, Anja Schulze: How to Support Knowledge Creation in New Product Development:: An Investigation of Knowledge Management Methods. *European Management Journal*, Volume 23, Issue 3, June 2005, pp. 263–273.
4. Ronald Maier, Thomas Hädrich: Centralized versus peer-to-peer knowledge management systems. *Knowledge and Process Management*, Special Issue: Mastering Knowledge in Organizations: Challenges, Practices and Prospects, Vol. 13, Issue 1, pp. 47–61, Jan/March 2006.
5. Peter van Baalen, et al.: Knowledge Sharing in an Emerging Network of Practice:: The Role of a Knowledge Portal. *European Management Journal*, Volume 23, Issue 3, June 2005, pp. 300–314.
6. Tiwana, A. and Bush, A.: Continuance in expertise sharing networks: A social perspective. *IEEE Transactions on Engineering Management* 52, 1 (Feb.2005), 85–101.
7. Kwang-Hoon Kim: A layered workflow knowledge Grid/P2P architecture and its models for future generation workflow systems. *Future Generation Computer Systems* 23 (2007) 304–316.
8. James S.H. Kwok, S. Gao: Knowledge sharing community in P2P network: a study of motivational perspective. *Journal of Knowledge Management*, 2000, Vol. 8 Iss: pp. 94–102.
9. Peter Haase, Ronny Siebes, Frank van Harmelen: Expertise-based peer selection in Peer-to-Peer networks. *Knowledge and Information Systems*, April 2008, Volume 15, Issue 1, pp. 75–107.
10. Atul Nerkar, Srikant Paruchuri: Evolution of R&D Capabilities: The Role of Knowledge Networks Within a Firm. *Management Science* (2005), Volume 51 Issue 5, May 2005, pp. 771–785.
11. Wai Fong Boh: Reuse of knowledge assets from repositories: A mixed methods study. *Information & Management* 45 (2008) 365–375.
12. Sheng-Tun Li, Won-Chen Chang: xploiting and transferring presentational knowledge assets in R&D organizations. *Expert Systems with Applications*, Volume 36, Issue 1, January 2009, pp. 766–777.

A Survey on Big Data Architectures and Standard Bodies

B.N. Supriya, S. Prakash and C.B. Akki

Abstract Huge amount of data is created due to the advancement in communications, digital sensors, computation and storage, business information, science, government, and society. It is expected that about 4 zettabytes of electronic data are being generated per year. Big Data is a term applied to data sets whose size is beyond the ability of available tools to undertake their acquisition access, analytics in a reasonable amount of time. The goal of this paper is to do a survey on various architectures available/proposed in the literature to meet the big data requirements. The features of these proposed architectures have also been discussed in the paper. The international standard bodies established for the development of big data domain is explored.

Keywords Big data architecture • Analytics

1 Introduction

Massive digital data is now a fact of life. Unknowingly, we have entered an era of Big Data which is a technology that provides the platform for automation of volume, variety, velocity, veracity, and many more dimensions. This has prompted us to explore more on the architecture of Big Data. As a result, this paper has been written to start with.

B.N. Supriya (✉) • C.B. Akki

Department of Information Science & Engineering, SJB Institute of Technology,
Bangalore 560060, India
e-mail: bnsupriya@gmail.com

C.B. Akki

e-mail: akki.channappa@gmail.com

S. Prakash

Department of Computer Science & Engineering, Dayananda Sagar University,
Bangalore 560078, India
e-mail: prakash.hospet@gmail.com

This paper is organized as follows: In Sect. 1, historical growth, definition, and applications of big data are discussed. In Sect. 2, the different architectural requirements have been listed. Section 3 deals with some architecture proposed by different vendors. A discussion on the list of standard bodies working for big data is given in Sect. 4. In Sect. 5, the summary of proposed architectures available in literature has been discussed. Finally, conclusions are discussed in Sect. 6.

1.1 History

We observe major technological developments in computer technology, internet technology and telecom domains for the last one century. Recently computer technology and internet technology are being used as a tool for many other technologies. Big data is one such technology that has emerged due to the development of digital electronics, computer technology and cloud computing.

People started feeling the over load of information as early as sixteenth century [1]. During first half of twentieth century, Father of “information age” Mr. Shannon could estimate the order of magnitude of largest information that he could think during his time [2]. Later with the dawn of the Internet, many technologies that produce large quantum of data started evolving. The phrase, “Information Explosion,” was first time used to quantify the growth of volume of data in Oxford dictionary as early as 1941 [3]. In 1944, a book entitled “The scholar and the Future of the Research Library” was published by Fremont Rider which estimated that in every 16 years volume of books in American University Libraries was doubling [3]. Further many researchers like B.A. Marron, Arthur Miller, I.A. Tjomsland, Peter J Denning, R.J.T. Morris have published “The information explosion”, “The Assault on privacy”, “Where do we Go from Here?”, “Tracking the Flow Of Information” and “Saving All the Bits” which gave a way to data tsunami. In 1998, John R Masey presents a paper “Big data ... and the Next Wave of Infrastrass” which establishes some of the challenges of big data and tries to overcome them. He was the first person who coined the term “Big Data” [3]. Some of the domains like Telecom, Living Environment, Social media and networks, etc., are major disciplines for the upraising of big data.

1.2 Definition

One can find more than 30 definitions from the literature which are given by the researchers in the domain. Doug Laney defines Big Data in terms of the three Vs: Volume, Velocity, and Variety [4]. This is the most venerable and well-known definition, first coined by Doug Laney. Gartner defines Big data as high volume, velocity and variety of information assets that demand cost-effective, innovative forms of processing for enhanced insight and decision making [5]. Rouse [6]

defines Big data as the voluminous amount of unstructured and semi-structured data that a company creates. This data would take too much time and cost for analysis. Although big data doesn't refer to any specific quantity, the term is often used when speaking about Petabytes and Exabyte's of data. In simplest form, the information that can't be processed or analyzed using traditional processes or tools are termed as big data.

1.3 Applications

Big data has a vast number of applications from underground physics experiment to Global positioning system. Big data helps the government to keep track of massive number of different archives and records of the country, census data, information about the people and improvements in the country [7–9]. In persistent surveillance sensors, processing of large data in parallel and in near real time is achieved with the help of big data [10, 11]. In health domain, big data can play a major role by supporting information retrieval methods to identify relevant disease diagnosis [12–14]. With the help of big data, mobile phones are becoming much smarter. Technically, more apps are getting developed which are making mobile phones more user friendly and more advanced. Now this available voluminous data is being used for making business decisions for enhancing profitability. A separate field known as Business Intelligence (BI) is becoming popular. Hierarchical Learning, Social Media, Ecosystems Research, Astronomy and Physics, Earth, Environmental and Polar Science, etc., are some of the disciplines using big data extensively. These applications add operational complexities in big data which is discussed in next section.

2 Architectural Requirements

Big data has different characteristics such as volume, velocity, variety, veracity. Depending on these characteristics, big data has about 25+ architectural requirements as per NIST [15]. Major five requirements are listed below:

1. Transferring of high volume of data to remote batch processing system.
2. System has to support both standstill and moving data.
3. Robust attribution defining complex machine/human processing.
4. Has to support both batch and real time processes.
5. The processed data has to be presented as per the user requirement.

3 Architectures

Research teams across the globe have proposed different architectures so as to meet the requirements. There are about 25 plus architectures available in the literature. Only few of them are given here as examples:

3.1 Architecture 01 [16]

This architecture consists of budding data models and required infrastructure. Figure 1 is a data-centric architecture that deals with data flow and data transformation. It mainly exposes projected “interoperability surface” and helps in identifying security and privacy issues. Some of the examples where this architecture has been implemented are in advertising agency, enterprise data warehouse, etc.

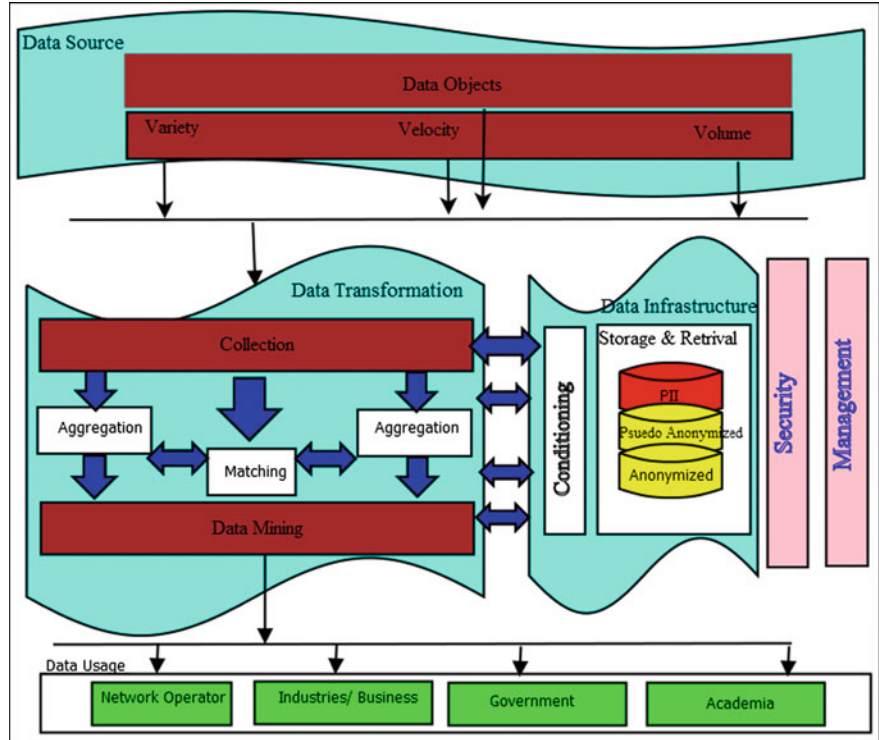


Fig. 1 Big data ecosystem reference architecture (Source [16])

- The key components and the description of Fig. 1 are given as below
1. Data Source: The data is collected by different sources and are classified into three categories: Volume, Velocity, and Variety.
 2. Data Transformation: The useful patterns are extracted from the processed data in various ways. Matching, collection, aggregation, data mining are some of the functions used for transformation of data.
 3. Data Infrastructure: In this, for the purposes of data transformation collection of data storage, servers and networking are used. It is sited to the right of the data transformation to highlight the natural role of the infrastructure.
 4. Data Usage: The different format, granularity and security are provided to the users from the processed data.

3.2 Architecture 03 [17]

LexisNexis introduced a platform named as High Performance Computing Cluster (HPCC) system which is mainly knob massive, multi-structured dataset. The architecture shown in Fig. 2 is supported on distributed architecture and meets all the requirements of data intensive computing applications.

The architecture mainly implements two distinct cluster processing environment—one for ‘data refinery’ known as ‘Thor’, the other for data delivery known as ‘Roxie’.

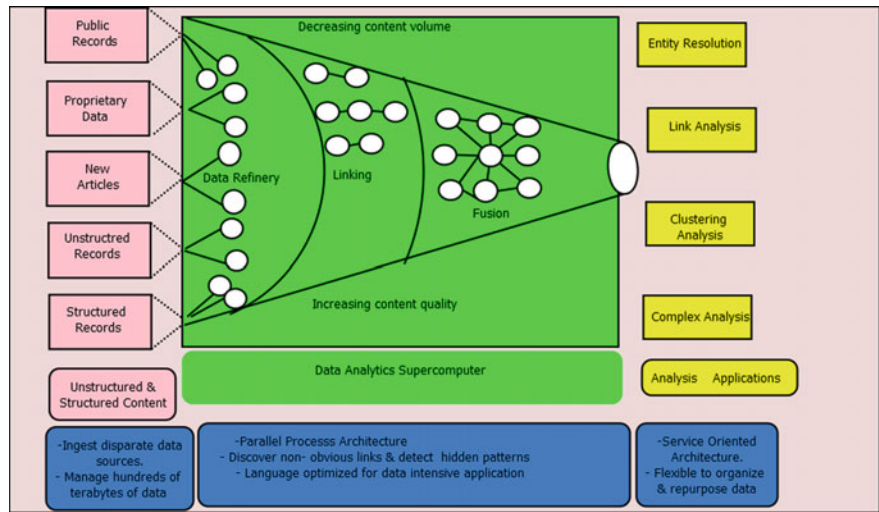


Fig. 2 LexisNexis vision for a data analytics supercomputer (Source [17])

- **Data refinery:** It is an ETL engine which is used for performing tasks like join, merge, sort, transform, etc.
- **Data delivery:** It serves as a parallel high throughput, structured query response engine that is suitable for performing volumes of structured queries. This can be seen as the data analytics supercomputer where the data is linked and fused to get the required output.
- **Enterprise Control Language (ECL):** An open source data-centric programming language that is used by both Thor and Roxie for data management and query processing.

The big data architecture designed by Oracle provides an entire vision of allied technical ability, how they integrate and fit together into large environment. This architecture provides a solution for enterprise big data requirement [18]. The big data architecture developed by SAP has introduced HANA which is optimized for both transactional and analytical processing. HANA when incorporated with Hadoop enables customers to transact the data between Hive, HDFS, HANA, or SAP Sybase IQ Server [19]. Architecture was proposed by Bob Marcus has been implemented by Apache. The architecture recognizes, lines up the system abilities, facilitates, and aligns the requirement. It is designed to support the big data requirements, use cases and technology gaps and uses layered model [20]. From the review we can accomplish that the architectures revolve around four main blocks: source, storage, analysis, and the end users.

Even though we have 25 plus architectures available in the literature one may not find the standard architecture that can be employed for all the objectives. A few standard bodies are working currently on the Big Data Architecture. The discussion of these standard bodies follows next.

4 Standard Bodies

4.1 *National Institute of Standards and Technology (NIST) [21]*

It was started in 1901. It is a measurement standards laboratory that is a non regulatory agency of the United States Department of Commerce. The nerve center is in Maryland and operates in Colorado. NIST helps in development of big data technology roadmap which defines and prioritizes requirements for analytics, usage, and technology infrastructure in order to support effective and secure adaption of big data. The group that works for big data is called as NIST Big Data Working Group (NBD-WD).

4.2 Open Data Center Alliance (ODCA) [22]

It is an independent organization which was launched in October 2010 which is now working on the standards for Big Data. The ODCA mainly works on the standards for cloud computing. They provide the conceptual model and also show the roadmap to the modern techniques.

4.3 Tele Management Forum and The Network Management Forum (TMF) [23]

It was established in 1988. In telecommunications and entertainment industries, TMF is acts as a nonprofit association, for service providers and suppliers. They provide the core framework which helps in business metrics.

4.4 Resource Description and Access (RDA) [24]

Resource Description and Access (RDA) is a standard for cataloging which provides instructions and guidelines on formulating data for resource description and discovery. RDA was initially released in June 2010. In March 2012, Library of Congress anticipated the implementation of RDA cataloging to be completed by March 31, 2013.

5 Summary of the Big Data Architecture

From the study on different architectures, we can summarize the interrelationship between architecture and general architectural requirements (discussed in Sect. 2) in the form of the table (Table 1).

Table 1 Interrelationship between architecture and architectural requirement

Requirements	1	2	3	4	5
Architectures					
1	+	+			
2	+	+		+	+
3	+	+	+	+	+
4	+	+		+	+
5	+	+	+	+	+

With the reviews on all these architectures we can design our proposed architecture as consisting of following major components:

5.1 Data Source

The data which comes from the heterogeneous source like Enterprise Legacy system, Data Management System, Smart Devices, etc., can be structured, semi-structured, unstructured. These data can vary in format and origin. These data are gathered to a file system called Landing zone. Further, these files are segregated into subdirectories based on the data type. Any updations on the files like naming and extension can be done in this layer.

5.2 Data Messaging and Storage Layer

In this layer, the collected data is segregated and loaded based on the metadata and prepared for transformation which is done with different components such as:

Data Acquisition: Acquires data from various data sources and sends to data digest component. This component must be able to determine whether the data should be messaged before it can be stored or the data can be directly sent to the analysis layer.

Data Digest: This component is responsible for messaging the data in the format required to achieve the purpose of the analysis. The loaded data is broken down into smaller chunks of files. A catalog of files is prepared and the corresponding metadata is processed. In this stage based on the user and processing requirements, data can be partitioned either horizontally or vertically.

Distributed Data Storage: This component is responsible for storing the data from data sources. Often multiple data storage options are available in this layer such as Distributed File Storage (DFS), Cloud, Structured Data Sources, and NOSQL.

5.3 Analysis Layer

Based on the required business rules, the data is transformed. This layer has multiple components to execute which are discussed below:

Entity identification: This component is responsible for identifying and populating the contextual entities. The data digest component should complement this entity identification component by messaging the data into required format.

Analysis engine: This component can have various workflows, algorithms and tools that support parallel processing.

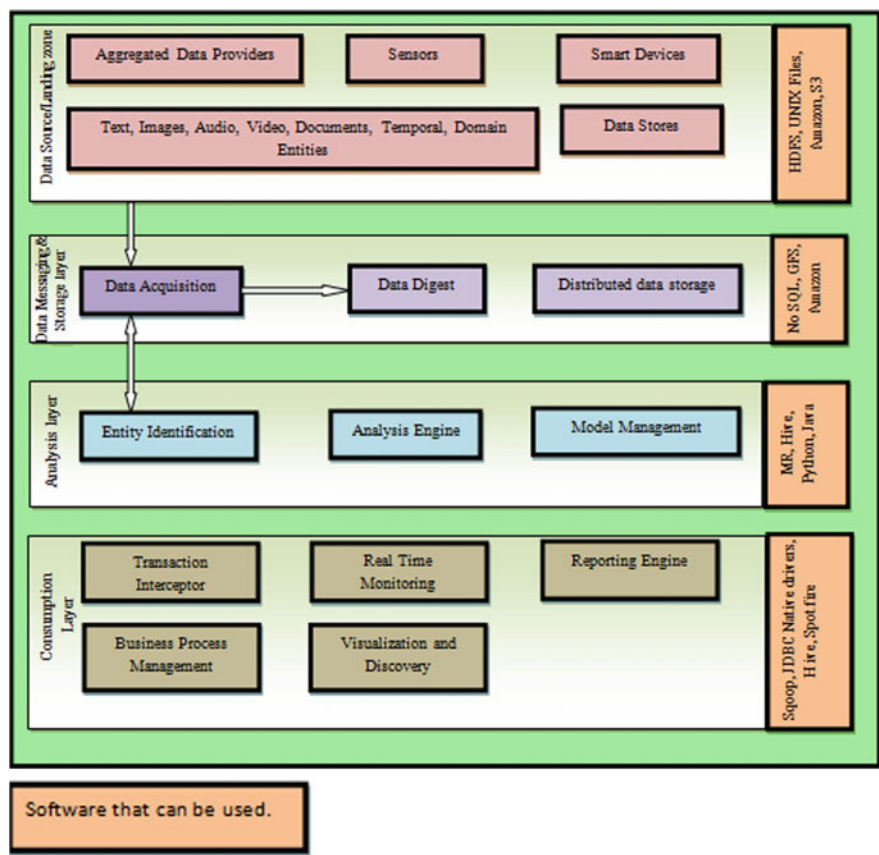


Fig. 3 Big data reference architecture

Model Management: This component is responsible for maintaining, verifying and validating the various statistical models by training them to be more accurate.

5.4 Consumption Layer

In this layer the resultant data set can be used for further processing like analysis, reporting, warehousing, integration, and visualization. All these layers can be depicted from the general architecture shown below (Fig. 3):

Other than these main processing layers, big data can also have some of the protocol defining layers such as security, privacy, data governance, etc.

6 Conclusion

As the amount of unstructured data grows, managing that data needs a new approach. Through better analysis, there is the potential for making faster advances in many scientific disciplines and resulting in profitability and success. In this paper, we have explored some major architecture and also the standard bodies proposed for big data. Looking into all these, we have proposed the architecture that satisfies the basic requirements.

References

1. Blair, Ann. 2003. Reading strategies for coping with information overload, ca.1550–1700. *Journal of the History of Ideas* 64, no. 1: 11–28.
2. Martin Hilbert, 2012, 1042–1055, *International Journal of Communication*, “How to Measure “How Much Information”? Theoretical, Methodological, and Statistical Challenges for the Social Sciences”.
3. <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/>.
4. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
5. Big Data definition, Gartner, Inc. [online] <http://www.gartner.com/it-glossary/big-data/>.
6. <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>.
7. <http://bigdatawg.nist.gov/usecases.php/M0147>.
8. <http://bigdatawg.nist.gov/usecases.php/M0148>.
9. <http://bigdatawg.nist.gov/usecases.php/M0219>.
10. <http://www.militaryaerospace.com/topics/m/video/79088650/persistent-surveillance-relies-on-extracting-relevant-data-points-and-connecting-the-dots.htm>.
11. http://stids.c4i.gmu.edu/papers/STIDSPapers/STIDS2012_T14_SmithEtAl_HorizontalIntegrationOfWarFig.hterIntel.pdf.
12. Regenstrief Institute, <http://www.regenstrief.org>.
13. <http://www.loinc.org>.
14. <http://www.ihie.org>.
15. <http://dsc.soic.indiana.edu/publications/NISTUseCase.pdf>.
16. Big Data Ecosystem Reference Architecture Orit Levin, Microsoft July 18th, 2013.
17. Anthony M. Middleton HPCC Systems: Introduction to HPCC (High-Performance Computing Cluster) May 24, 2011.
18. An Oracle White Paper in Enterprise Architecture—Information Architecture: An Architect’s Guide to Big Data August 2012.
19. <http://scn.sap.com/community/hana-in-memory/blog/2013/04/30/big-data-technologies-applications>.
20. Evaluation Criteria for Reference Architecture by Robert Marcus.
21. National Institute of Standards and Technology, <http://www.nist.gov/>.
22. http://en.wikipedia.org/wiki/Open_Data_Center_Alliance.
23. TeleManagement Forum and the Network Management Forum, <http://www.tmforum.org/>.
24. Resource Description and Access, <http://www.rdatoolkit.org/>.

Generating Data for Testing Community Detection Algorithms

Mini Singh Ahuja and Jatinder Singh

Abstract These days Internet usage has increased. People of all age groups use Internet and this has led to a new research field called complex networks. Complex networks such as social networks, biological networks, technological networks, etc., have become the interest of many researchers because of their wide range of applications. These complex networks have many properties like scale-free networks, transitivity, presence of community structure in these networks. Community detection is one of the most active fields in complex networks because it has many practical applications. In this paper we have studied about community detection. We have also discussed about the techniques of generating data for comparing various community detection algorithms.

Keywords Community · GN benchmark · LFR benchmark

1 Introduction

Nowadays real systems have grown in size tremendously. They contain millions of actors and have different relationships. Complex networks are the powerful modeling tools which represent most real-world systems. Complex network paradigm is one of the modeling tools which have spread through several application fields such as sociology, communication, computer science, biology, and physics, and so on during last decades. Complex networks can be represented in the form of large graphs which have large number of nodes and different types of relationships with nontrivial properties. These nodes can be anything: a person, an organization, a computer, or a biological cell. Nodes can have different sizes or attributes which

M.S. Ahuja (✉)

Guru Nanak Dev University, Regional Campus, Gurdaspur, India

e-mail: minianhadh@yahoo.co.in

Jatinder Singh

KC Group of Institutes, Nawashahr, India

e-mail: bal_jatinder@rediffmail.com

represent a property of real system objects. These graphs can be directed, undirected, or weighted. A complex network has its roots in graph theory. Few examples of complex networks are Internet maps (IP, Routers [1], web graphs (hyperlinks between pages) [2], data exchange (emails) [3], social networks (Facebook, Twitter, scientist collaboration networks), biological networks (protein interaction, epidemic networks), etc. Complex networks have nontrivial properties so they cannot be explained by uniform random, regular, or complete models [4]. This has resulted in definition of set of statistics which have become fundamental properties of complex networks. These properties are now being used by many researchers for studying various phenomena's like spreading of information [5], protocol performance, etc. But a major challenge in the study of complex networks is how to collect data for analysis. We cannot directly collect data from these real world complex networks to study them. So researches have to make an assumption that initially data is not fit to find the real properties but as the size of the data grows the properties become more and more stable. The research is going on this side of complex network too [3]. They are trying to find the impact of the measured procedures on the obtained data to study the induced bias [6].

2 Communities

Large real-world networks are generally characterized by heterogeneous structures which have some particular properties. The heterogeneous distribution of the links has led to community structure [7–9] A community is a set of entities which are linked to all the other entities in the network. The entities in one community perform same function and share some common properties. A community structure reveals the internal organization of the nodes. Different communities combine to form a complex network. In other words, a community can be described as a collection of vertices within graph which are densely connected among themselves but are loosely connected to the rest of the graph [10]. Communities can also be called as clusters or modules which share common properties. These communities have many features. They can have hierarchal or overlapping structure inside them. Moreover, these communities can be dynamic which change with time or can be multirelational (multiple relations). Many real networks such as social networks, biological networks exhibit community structure. This property of complex networks can be used in various applications such as to study the spread of disease in social networks [11]. Web clients who have similar interests and are geographically near to each other can be clustered to improve the performance of the service providers on the World Wide Web. Each cluster can be served by a dedicated mirror server. Community structure property reduces very large graph into smaller ones. Nowadays, community detection has become a popular field of research. Community detection algorithms are the common and fundamental tools which help to uncover the principles present in networks. The main aim of a community detection algorithm is to divide nodes or vertices of a network into any number of

communities or groups, maximize the number of edges between groups, and minimize the number of edges between vertices in different groups. Community detection algorithms focus only toward the network structure. While detecting the communities, two possible sources of information are expected: network structure, and the attributes and features of the nodes. Many algorithms for community detection have come up till now.

3 Different Definitions of Community Detection Algorithms

Community discovery problem is very similar to clustering problem of data mining. Community discovery is a clustering task of data mining which is done on the graphs. Till date many algorithms have been proposed by researchers which have different definitions.

Density-based algorithms These algorithms are based on the topology of the network edges. According to density-based algorithms, community is a group in which there are many edges between vertices but there are fewer edges between groups. These algorithms divide the network into groups which have maximum number of edges in each group and minimum number of edges between the groups.

Node similarity-based algorithms These algorithms define community as a group of nodes which are similar to each other but different from rest of the network. Similarity can be structural similarity, shortest path between nodes or location-based similarity (topological information, nodal attributes define location similarity).

Pattern-based algorithms These algorithms try to identify the largest pattern (cliques) with large common nodes. These algorithms show better performance than density-based algorithms as these do not rely only on numeric values.

Link Centrality-based algorithms Link centrality is based on two main features: number of nodes the link is connecting and how likely these connections are to be used. Link between communities are very central and few. So they are likely to be used mostly. Newman [12] defined edge betweenness measure by considering the total number of shortest paths going through a link. Radicchi et al. proposed edge centrality measure. It is defined as the ratio of the number of existing cycles containing the link of interest to the number of possible cycles given the existing links.

Other Algorithms Many authors have used data compression technique [13] and considered community as a set of regularities in the network topology which can be used to represent the whole network in a better way. A community founded by these algorithms will have maximum compactness and minimum information loss.

4 Literature Survey

The data needed for comparing community detection algorithms can be extracted by different methodologies. Many researchers have been working in this field.

Christopher Olston and Marc Najork [14] presented the basics of web crawling. In this paper, they discussed the crawling architecture and also gave information about the future scope of crawling. They have also elaborated on how the undesirable content can be avoided and also discusses the future directions in this field.

Raja Iswary, Keshab Nath [15], discusses the different techniques to develop a crawler and how to build an efficient crawler. They also elaborate on different crawling techniques like focused crawler, distributed crawler, incremental crawler and hidden web crawler. Also, the different design issues have been discussed in their paper. Malhotra [16], elaborates on the architecture of the web crawler and the different web crawling policies.

Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi [17], introduces a class of networks that explains the heterogeneity in the distribution of node degrees and community sizes. Lancichinetti and Fortunato [18], further continued their study and tested their benchmark on directed and unweighted graphs. They have also paid attention to overlapping communities which is an important characteristic of community structure in real world networks.

Jaewon Yang and Jure Leskovec [19], proposed the concept of ground truth communities which provides interesting future directions. Coscia et al. [20], organizes the different categories of community discovery methods based on the definition of community adopted by them.

5 Community Detection Problem

Detecting clusters or communities in real-world network is a problem of considerable practical interest. The community detection problem has plenty of challenges as it is highly related to the problem of clustering large heterogeneous datasets. Till date many researchers have proposed number of algorithms, but all the community detection algorithms are different from each other and are not clearly defined [21, 22]. So heterogeneity of different algorithms poses a challenge to community detection. Different networks (biological, social, etc.) have their own properties. This difference in properties as led to the unsolved question: which algorithm is suitable for which type of network?

Moreover, these algorithms do not detect the same communities. So the problem is how to compare the performance of these algorithms. Actually, the researchers are interested in following information.

- What type of information is used by the algorithm? A network can have different type of data: link attributes (weights, directions) node attributes, different types of links.
- What type of community produced (partition, overlapped). The nature of communities the algorithm identifies.

6 Testing of Community Detection Algorithms

In order to test the algorithms which are used for detecting communities of different complex networks, we need to extract the data from these networks. Community detection algorithms can be tested on following type of data:

- Data from real-world networks.
- Data from artificial networks and benchmarks.
- Data with ground truth communities.

6.1 *Data from Real-World Networks*

Networks are present in each and every aspect of our lives. We are surrounded by a number of networks like WWW (World Wide Web) is a network which we use everyday. The friendship between individuals, the business relations, etc., are all networks.

It is very difficult to test the different algorithms using real-world data. It is very costly and time consuming to obtain real-world data. Moreover complex networks have many properties such as average degree, shortest path, degree distribution, etc., which are very difficult to be controlled in real world networks. Real-world data can be downloaded from these networks by the use of web crawlers. Web crawler is software for downloading pages from the Web automatically. It is also called web spider or web robot. Many researchers have used web crawlers to get the web data for their research work. Gjoka [23], had used web data (extracted by web crawler) to measure the statistical properties of online social networks. Catanese et al. [24], has used crawlers on the social networking site, i.e., Facebook. This data was later used to study the community structure of face book.

6.2 *Ground Truth*

Ground Truth communities are explicitly labeled functional communities. It is nearly impossible task to find explicitly labeled communities. If we are able to find

the ground truth communities then we can link the structural definition of a network with the functional definition of the network. The structural definition of network community is based on the structure of the connectivity between a set of nodes while the functional definition is based on the common function or role that the community members share. Social networks found interest-based communities like students of same school; people interested in singing career, etc.

6.3 Artificial Networks and Benchmarks

Artificial networks provide solution to all these problems. They are widely used to compare the performance of different community detection algorithms. We can easily generate artificial networks with desired properties using generative models. But these cannot be substitute to real-world data; instead can act as complement.

6.3.1 Girvan And Neuman Benchmark

The first benchmark for testing these algorithms was developed by Girvan and Newman called as GN benchmarks. GN benchmark is very simple to use. This benchmark data set contains 128 vertices which are divided into 4 groups of 32 nodes each. Further each vertex has a degree of 16. This benchmark can detect only disjoint communities as here each vertex is associated with only one community during network generation. The strength of each community association depends on the mixing parameter μ which gives the probability of where the edge will be placed.

Mixing parameter is given by:

$$\mu = k_o/k_i + k_o$$

k_o is number of edges connecting a vertex to a vertex in another community.

k_i is the number of edges connected to a vertex.

Many algorithms give good result with GN benchmarks as all communities identified by them are identical in size. GN benchmarks produce networks with Poisson distribution but real-world networks follow power law distribution. Moreover, sometimes these benchmarks fail to detect communities because of the fluctuations in distribution of links. So GN benchmarks are not so fruitful in comparing community detection algorithms.

6.3.2 LFR Benchmark

There are several conditions that need to be considered in GN benchmarks:

1. All nodes of the network have essentially the same degree.
2. The communities are all of the same size.
3. The network is small.
4. Have Poisson degree distribution.

These are generally the drawbacks of this benchmark that prevent their use in the real networks. Real networks have a heterogeneous distribution of all the nodes. For a benchmark to be reliable in real networks, it should consider the communities of very different sizes. Due to these reasons LFR benchmarks proposed by Lancichinetti et al. have replaced the GN benchmarks [17]. LFR benchmarks give more realistic model of a real world graphs. It is a special case of planted l -partition model. These benchmarks can generate undirected and unweighted networks with mutually exclusive communities. These benchmarks can also detect overlapped communities. With these benchmarks we can create scale free networks with communities of varying sizes. The advantages of using LFR benchmarks are:

1. LFR benchmarks can work for higher value of μ (mixing parameter) as power law distribution is used for node degree distribution and community size.
2. LFR benchmarks are better in showing the reliability of a community detection algorithm for real applications.

After generating synthetic networks with any network generator, the accuracy of community detection algorithm can be determined by comparing the discovered community with the ground truth.

More over, there are many metrics to measure the quality of the communities detected by these algorithms. One popular metric is modularity [10]. Modularity measure has been used by many authors in their research to compare the various community detection algorithms. Few others are Rand Index (RI), Purity, Normalized mutual information (NMI) and F measure. Many researchers have even compared these measures to compare which one gives the best result for all community detection algorithms.

Community detection in complex networks is a very challenging problem. Much work as been done in this field but still it is not clear which algorithm to be used in what situation.

7 Conclusion

A complex network is a very young and promising interdisciplinary field whose roots lie in graph theory. The field of complex networks is helpful in understanding many complex phenomena's such as spam detection, protein interaction, spread of disease, etc. Complex networks have many properties which have been studied by

many authors in the past. Community detection is one of the fields of complex network which has gained a lot of attention in today's world. Many algorithms have been proposed by different researchers but still many questions are unsolved.

References

1. C. Gkantsidis, M. Mihail, and E. Zegura, "Spectral analysis of internet topologies," in *IEEE INFOCOM*, 2003.
2. P. Boldi and S. Vigna, "The webgraph framework i: compression techniques," in *WWW*, 2004.
3. S. Voulgaris, A.-M. Kermarrec, L. Massoulié, and M. van Steen, "Exploiting semantic proximity in peer-to-peer content searching," in *IEEE FTDCS*, 2004.
4. R. Albert and A. Barabási, Statistical mechanics of complex networks, *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.
5. Pastor-Satorras, R. and A. Vespignani. Epidemic spreading in scale-free networks [Journal Article]. *Physical review letters*, 2001. 86(14): p. 3200–3203.
6. Z. Li and P. Mohapatra, "Impact of topology on overlay routing service," in *IEEE INFOCOM*, 2004.
7. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, 2010.
8. Richardt J and Bornholdt S, 2006, Statistical mechanics of community detection, *Physical Review E* 74, 016110.
9. Gulbahce, N. and S. Lehmann. The art of community detection [Journal Article]. *Bioessays*, 2008. 30(10): p. 934–938.
10. M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, 2004.
11. Goh K-II, Cusick ME, Valle D, Childs B, Vidal M, et al. 2007. The human disease network. *Proc Natl Acad Sci USA*.
12. M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, January 2005.
13. Rosvall M and Bergstrom C T 2007 An information-theoretic framework for resolving community structure in complex networks *Proceedings of the National Academy of Sciences of the United States of America* 104 7327–31.
14. Christopher Olston and Marc Najork (2010) "Web Crawling", now the essence of knowledge, Vol. 4, No. 3 (2010) 175–246.
15. Raja Iswary, Keshab Nath (2013), "Web Crawler", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, Issue 10, ISSN: 2278–1021.
16. Mohit Malhotra (2013), "Web Crawler And It's Concepts.
17. Lancichinetti A, Fortunato S and Radicchi F 2008 Benchmark graphs for testing community detection algorithms *Phys Rev E* 78 046110/
18. Andrea Lancichinetti and Santo Fortunato (2009), "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities", *Physical Review E* 80, 016118/
19. Jaewon Yang and Jure Leskovec (2013), "Defining and Evaluating Network Communities Based on Ground-Truth", Springer, doi: [10.1007/s10115-013-0693-z](https://doi.org/10.1007/s10115-013-0693-z).
20. Michele Coscia, Fosca Giannotti and Dino Pedreschi (2010), "A Classification for Community Discovery Methods in Complex Networks", *Wiley Online Library*, Volume 4, doi: [10.1002/sam.10133](https://doi.org/10.1002/sam.10133).
21. Andrea Lancichinetti, Santo Fortunato, 2009, Community detection algorithms: a comparative analysis, arXiv: 0908.1062v1 physics soc-ph.

22. Hubert L and Arabie P 1985 Comparing partitions *Journal of Classification* 2 193–218.
23. Minas gjoka (2010), “Measurement of Online Social Networks”, university of california, Irvine.
24. Salvatore A. Catanese, Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, Alessandro Proveti (2011), “Crawling Facebook for Social Network Analysis Purposes”, WIMS’11, May 25–27, 2011 Sogndal, Norway, ACM 978-1-4503-0148-0/11/05.

Metamorphic Malware Detection Using LLVM IR and Hidden Markov Model

Ginika Mahajan and Raja

Abstract This paper proposes a new method to detect metamorphic malware with the help of hidden Markov model and LLVM intermediate representation. The new approach improves the accuracy of HMM by simplifying various uncertain transformations present in the metamorphic code with the help of conversion of these instructions into the LLVM IR. Due to conversion of the unstructured assembly language code into the simplified LLVM IR, many of the code obfuscations are reversed and thus simplified form of instructions are generated. We can easily detect the remaining transformations or other unknown probabilistic states which HMM undergoes. Conversion to LLVM IR increases the predictability of HMM and also the probability to successfully detect other hidden states of malwares. Hence, this approach to first convert code into IR and then test the IR on HMM increases the probability of successful detection of metamorphic malwares.

Keywords Code obfuscation · LLVM intermediate representation · HMM · Bit code

1 Introduction

Metamorphic malwares are expert in mutating their code [1, 2] and transforming them from one structure to another without any impact on their functionality [3, 4]. These malwares are very hard to detect due to lack of proper syntactic signature for each variation that makes them almost invisible for detection by the detectors. This phenomenon of changing signatures continuously can make millions of signatures for single malware and it is really impossible to store infinite number of signatures for the single malware. Due to this property of changing signatures after each

Ginika Mahajan (✉) · Raja
Central University of Rajasthan, Ajmer, Rajasthan, India
e-mail: ginikamahajan@curaj.ac.in

Raja
e-mail: raja@rediffmail.com

infection, it is infeasible to make syntactic signature database even for a single malware.

There are several techniques to detect such kind of malwares based on machine learning [2]. Due to high content of mutations in their code, machine learning models lack sufficiently to get a good percentage of detection. The hidden Markov model has been widely used for detection of metamorphic malwares [3] but along with it the same researchers had developed the malware which could easily evade those techniques based on hidden Markov model. More efficient technique like tiered HMM approach was used to enhance the efficiency of the HMM to increase the detection rate [5]. HMM combined with chi-squared testing [6, 7] was used to detect the metamorphic malware and this proved successful but later same researchers created the malware which evaded their technique [7]. We need to see whether we could increase the predictability of HMM by reversing some of the obfuscations and then inputting the partially clean code to the HMM and getting the results. We use LLVM [5, 8, 9] to produce the partially clean code which is free from some of the obfuscations and then use the opcode sequences from this code, to see whether it can improve the predictability of HMM. LLVM will produce the LLVM bitcode [10] and this bitcode will then be used as an intermediate form to further optimize the code to that level where it is almost free from basic obfuscations. Afterwards this bitcode is transformed into the x86 executable at backend. This partially clean code will be used as the base code for the HMM model to test and the training code will be processed in the same way.

This paper is divided into following sections. The first section contains the basic obfuscations which are present in the code. In the next section, we propose our method and define its various phases. Then we explain some of the most relevant work done about the use of hidden Markov models to detect metamorphic malware. The final section consists of our future work and conclusion.

2 Code Transformations

Code obfuscations are the main techniques which are used to change the structure of code keeping the base functionality same. These transformations mutate the code syntactically but keep the semantics of code same. We discuss these obfuscations briefly in the following section.

2.1 *Nop Semantic Instructions*

Nop semantic instructions [11] are those instructions which are actually doing nothing. These instructions are just like the dead code and they either do not execute or even if they execute they exactly do nothing. These type of instructions are also used as delay instructions, because they just increase the delay [4, 12] but

actually have no functionality. These instructions are successfully used to mutate the code for a long time. As advanced processors neglect these instructions, nowadays, they are very less used in the malwares.

2.2 Dead Code

Dead code is the code or an instruction [13] which on execution produces no effect. It is exactly same as the nop semantic instruction. It is also called as junk code as it only increases the bulk. The dead code can be of many types. Take an example of register operation like `mov eax, eax`. This is a valid instruction but upon execution of this, there is no change in the register value it remains same so it is considered as dead. There are many types of other dead instruction like `add`, `xor`, `sub`, etc. Figure 1 shows dead and nop semantic instructions.

2.3 Register Reassignment

Register reassignment is renaming registers only after each mutation is done as a valid code mutation. It has been successfully used in the Regswap virus as basic obfuscation. In this technique, the registers are assigned at different locations. Figure 2 shows the two different versions of the Regswap virus where just the position of registers is changed and the code is same elsewhere.

2.4 Unreachable Code

Inserting unreachable code via unconditional jumps is the most efficient technique used nowadays in the metamorphic malware. The code which is between the jumps

Original Code	With Nop Semantic and Dead Code
<code>mov edx,0x3289</code> <code>add eax,ebx</code> <code>sub eax,eax</code> <code>add ecx,0x3311</code>	<code>mov edx,0x3289</code> <code>mov edx,edx</code> <code>nop</code> <code>add eax,ebx</code> <code>sub eax,0x0000</code> <code>sub eax,eax</code> <code>add ecx,0x3311</code> <code>add ecx,0x0000</code> <code>push esi</code> <code>pop esi</code>

Fig. 1 Code inserted with dead and nop semantic instructions

Version	Version 2
<pre>pop edx mov edi,0004h mov esi,ebp mov eax,000Ch add edx,0088h mov ebx,[edx] mov [esi+eax*4+00001118],ebx</pre>	<pre>pop eax mov ebx,0004h mov edx,ebp mov edi,000Ch add eax,0088h mov esi,[eax] mov [edx+edi*4+00001118],esi</pre>

Fig. 2 Two versions of Regswap virus

Original Code	With Unreachable Code
<pre>mov edx,0x3289 add eax,ebx sub eax,eax add ecx,0x3311</pre>	<pre>mov edx,0x3289 jmp label Junk code ... label add eax,ebx sub eax,eax add ecx,0x3311</pre>

Fig. 3 Unreachable instructions

statements may not execute. This code is unreachable and so never executes. The use of unconditional jump statements are widely used in modern malware as it never unveils until it is executed. Figure 3 shows how the jump goes directly to the label.

2.5 Instruction Reordering

This is simple code mutation where the instructions are reordered. Only the position of the instructions is changed. If the instructions are independent from each other then there is no problem in making permutations of the instructions while reordering. There can be instructions which are dependent upon each other; their use of unconditional jumps comes handy as the state for HMM. Figure 4 shows the instruction reordering obfuscation.

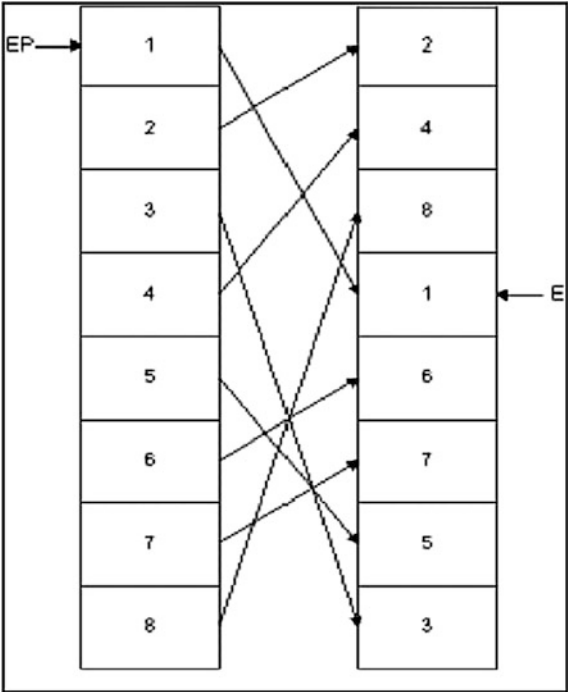
Original Code	Reordered Instructions
<code>mov ebx,0x7647</code> <code>push ecx</code> <code>pop eax</code> <code>add esi,[0x3452]</code>	<code>add esi,[0x3452]</code> <code>mov ebx,0x7647</code> <code>pop eax</code> <code>push ecx</code>

Fig. 4 Instruction reordering

2.6 Subroutine Reordering

Subroutine reordering is just to permute the main function and explicitly call them via jump statements. The subroutine reordering obfuscation is very hard to tackle statically because a malware having, say, 10 subroutines can produce 10! variants that means 3,628,800 versions of a single malware. These kind of obfuscations are dealt with emulations only by running them directly in virtual machine. Figure 5 shows the malware using subroutine reordering.

Fig. 5 Subroutine reordering



2.7 *Equivalent Instruction Substitution*

Many instructions semantically perform same task or some of the instructions combined with some other instructions can do the work of other instruction. The equivalent instruction substitution is the obfuscation method where the equivalent instructions are substituted in place of the other instructions. For example, the sub a, a can be replaced by xor a, a. Similarly, there are many other instructions which do the equivalent work. Figure 6 shows the equivalent instruction substitution obfuscation.

3 **Proposed Work**

We propose a method where the training data set and test data set is first optimized via LLVM optimizer passes. The optimizer passes remove almost all the dead code and nop semantic instructions. The constant propagation also removes the equivalent substituted instructions. The use of the optimized data as the training and test data has its own benefits which we have discussed further. We propose a method that tests the optimized malware code on the HMM (Fig. 7).

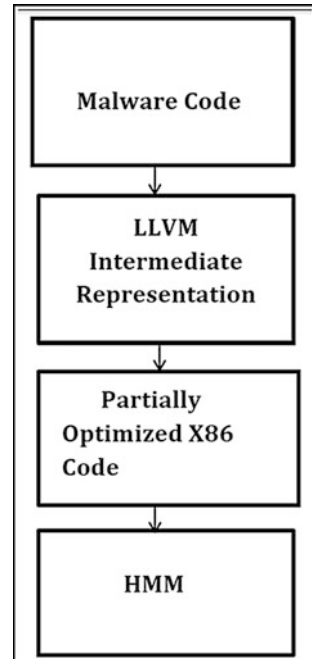
The algorithm of the proposed methodology is as follows:

- Step 1: Disassemble the executable with any disassembler like IDA Pro.
- Step 2: Extract the whole assembly code of the malware.
- Step 3: Convert this code into LLVM intermediate representation.
- Step 4: Run LLVM optimizer passes on this LLVM bitcode or LLVM IR.
- Step 5: Convert it back into x86 code.
- Step 6: Use this optimized x86 code as main test and training data.
- Step 7: Train HMM with this optimized data.
- Step 8: Test HMM with this optimized data.
- Step 9: Results.

Original Code	Equivalent Instruction Substitution
<code>mov edx,0x3289</code> <code>add eax,ebx</code> <code>sub eax,eax</code> <code>add ecx,0x3311</code>	<code>mov edx,0x3200</code> <code>add edx,0x0089</code> <code>xor eax,eax</code> <code>add ecx,0x0011</code> <code>add ecx,0x3300</code>

Fig. 6 Equivalent instruction substitution

Fig. 7 Overview of proposed method



3.1 Converting ASM to LLVM IR

First of all, we disassemble the executable by the interactive Disassembler Pro. The full code sequence is extracted from the IDA Pro which is then taken as input to the LLVM frontend. This x86 code is converted into the intermediate representation. To run the optimizer passes, we have to convert the assembly code into the LLVM IR. Converting the x86 code into LLVM IR is cumbersome. Since x86 executable code is unstructured, we need to develop the frontend for the LLVM that converts the x86 code into the intermediate representation. This conversion can be done manually or can be automatically converted into the IR. We do not have the LLVM frontend for the x86 executable till date and to write the frontend for it is very hard as the x86 code is very unstructured and simple instruction changes frequently. So, writing the frontend for the LLVM compiler infrastructure is quite hard and has not been done yet by any of the developers. Since, in our proposed method, we have to either do the conversion manually or write the frontend for the LLVM.

After this, x86 codes will be converted into the LLVM intermediate representation. LLVM IR is single statement assignment form which is ideal for the code optimization. The SSA form is ideal as it internally resolves all the conflicts which are present in the x86 code or any other assembly format. The resolution of the conflicts is necessary as it can hinder the optimization process and it mostly effects the equivalent instruction substitution obfuscation and constant propagation. The LLVM has its own instruction set and its own infinite number of virtual

registers. Having own instruction set has its own benefit of substituting the various machine instructions by its own instructions and the presence of the infinite number of registers is the main reason that the conflicts are easily removed in the LLVM IR. The virtual registers make it possible to store the conflicting values or the conflicting variables in the temporary registers and thereafter resolve them. These registers also help in the constant propagation by storing the value temporary and then using one of these values after they are used. Hence, its own instruction set and infinite number of registers makes it possible for us to easily simplify and reverse the basic obfuscations in the code.

3.2 Running LLVM Optimizer Passes

After generation of the LLVM IR we need to run the LLVM optimizer passes on the intermediate representation. These optimizer passes removes most of the obfuscations such as dead instruction removal, constant propagation, equivalent instruction substitution, and some of the unreachable code unless it is not defined by the unconditional jumps. The LLVM optimizer optimizes this intermediate code representation to the maximum extent. This code is now clean of almost all the basic code transformation or code obfuscations. This intermediate representation now contains the advanced code obfuscations if any has been introduced by the malware writer like the unconditional jumps, etc.

The advantage of running the optimizer passes on the intermediate language is that since it is in the single statement assignment form it easily reduces the conflicts and simply removes the dead code. The foremost important thing to run the optimization on the LLVM IR is that it substitutes equivalent instructions with the predefined common instructions. It efficiently does the constant propagation which is mandatory for the unnecessary instruction removal. After the IR is optimized by all the LLVM passes, this IR is changed back to the partially optimized x86 code which is used as the test and training data for the HMM models.

3.3 HMM Test and Training Data

The optimized x86 code from the LLVM can now be used as the test and training data. Since this optimized code is used as test data the obfuscations which are contained, are only the advanced obfuscations and are more predictable now. Reason being for basic dead code patterns, equivalent instruction obfuscations are no more present. The patterns on which the HMM is trained are only the advanced patterns which are present in the code. These patterns include the labeled

unconditional jumps and the HMM will easily train itself with these patterns. Other patterns are already removed so it will be very much obvious for the HMM that the patterns available will be of the similar kind that will remain after the code is freed from the basic obfuscations.

3.4 Optimized Code: A Better Data to Test and Train

Since HMM works on determining the unknown states as certain previous states are known. Here, we have reduced the number of unknown states in the code. These reductions were possible by converting the code into the LLVM intermediate representation. The optimization of the code will improve the predictability of the HMM as it can be initially made sure that these certain states will never occur in the test data and training data and certain states have high probability of being into the code. The states which have the minimal probability of being in the code are the states and the patterns which come due to the dead code insertion, equivalent instruction substitution, and nop semantic instruction insertion. And the states which have high probability of being in the code are the states due to the instruction reordering and unconditional jumps present in the code. Hence the predictability to detect the patterns can be enhanced if the trained data and test data to HMM are optimized. Thus, the optimized data is better for HMM to accurately predict the malware family.

4 Related Work

There are numerous techniques used to detect the metamorphic malware but most of them are either syntactic-based or behavioral-based analysis which may sometimes give many false positives. Techniques based on hidden Markov models have been proposed to detect such kind of malware. These techniques try to find the next state of the malware without having knowledge of the previous states. This method proved to be very good for some of the malwares but failed while detecting others. Enhanced version of the same technique has been proposed like tiered HMM and HMM and Chi-squared testing based on [4]. But the same authors developed the malware which could easily evade these techniques.

Graph-based solutions were given too, but with many false positive results. Some authors gave the solutions on counting the number of control flow graphs created by the malware and counting them. Classification was done on the basis of the number of control flow graphs in a malware. Methods based on eigenfaces and eigenvector were given by. These methods model the metamorphic malware as the eigenvirus just like the eigenface.

Rank linear discriminant analysis method was given by in which the authors used rank linear discriminant analysis to rank the opcode or opcode sequences and subsequently reduce that to the needed opcode sequence that provided to be very useful for the detection of certain particular malware and its variants.

Behavioral detection base on the capturing of critical API calls have been done by various authors. These methods are dynamic in nature and have proved very efficient in detecting with great accuracy. The register swap obfuscations were solved by the wildcard search and half byte scanning which were successful to very high extent but later on these methods showed some drawbacks. Gerardo et al. [8] have proposed a method which is based on number of push and pop instructions used by malware. The instructions used by one kind of malware and its variants are different from as used by other kinds of malware and its variants and as well form the benign code. Closely related work has been done by industry. Advanced Malware Laboratory, COSEINC, Singapore [14, 15] has done work. They have given OptiSig solution which used LLVM IR as intermediate form and conversion of the intermediate representation in Boolean logic. We mainly do not aim to do the work semantically as they have done. Our work is based on machine learning.

5 Future Work

In future, we will try to implement this work. To implement this work, there is a need to build LLVM frontend for x86 executable which will convert the x86 codes to LLVM IR or LLVM bitcode which will be subsequently optimized by the LLVM optimizer passes. We will test and train this model with the optimized x86 code which will be cleaned of all the basic obfuscations and will be more predictable than the earlier obfuscated code.

6 Conclusion

In the paper, we have proposed the method which will increase the predictability of the hidden Markov models in case of detecting the metamorphic malware. In this paper, we have proposed the model which uses partially clean training and test data. To predict the hidden states on the partially clean training and test data will be very useful in correctly predicting to greater accuracy. The use of hidden Markov models is very prevalent in detecting the metamorphic malware but we used the same technique not directly on the malware code but on the optimized form of the code. Due to optimization of code their might be less hidden states and due to this we may increase the predictability to greater accuracies.

References

1. Zhang, Qinghua, and Reeves, D.: Metaaware: Identifying metamorphic malware. In: Twenty-Third Annual IEEE Computer Security Applications Conference, ACSAC (2007).
2. Walenstein, Andrew, et al.: Normalizing metamorphic malware using term rewriting. In: Sixth IEEE International Workshop on Source Code Analysis and Manipulation SCAM'06 (2006).
3. Attaluri, Srilatha, Scott McGhee, S., and Stamp, M.: Profile hidden Markov models and metamorphic virus detection. *Journal in computer virology* 5.2, 151–169 (2009).
4. Toderici, Annie, H., and Stamp, M.: Chi-squared distance and metamorphic virus detection. *Journal of Computer Virology and Hacking Techniques* 9.1, 1–14 (2013).
5. Lattner, Chris, and Adve, V.: The LLVM compiler framework and infrastructure tutorial. *Languages and Compilers for High Performance Computing*. Springer Berlin Heidelberg, 15–16 (2005).
6. Lin, Da, and Stamp, M.: Hunting for undetectable metamorphic viruses. *Journal in computer virology* 7.3, 201–214 (2011).
7. Govindaraju, Aditya.: Exhaustive statistical analysis for detection of metamorphic malware. (2010).
8. Shanmugam, Gayathri, Low, R., and Stamp, M.: Simple substitution distance and metamorphic detection. *Journal of Computer Virology and Hacking Techniques* 9.3, 159–170 (2013).
9. Lattner, Arthur, C.: LLVM: An infrastructure for multi-stage optimization. Diss. University of Illinois at Urbana-Champaign (2002).
10. Lattner, Chris, and Adve, V.: LLVM: A compilation framework for lifelong program analysis transformation. In: IEEE International Symposium on Code Generation and Optimization, CGO (2004).
11. Baysa, Donabelle, Low, R., and Stamp, M.: Structural entropy and metamorphic malware. *Journal of Computer Virology and Hacking Techniques* 9.4, 179–192 (2013).
12. Sridhara, Madenur, S. and Stamp, M.: Metamorphic worm that carries its own morphing engine. *Journal of Computer Virology and Hacking Techniques* 9.2, 49–58 (2013).
13. Chouchane, Mohamed R., and Lakhotia, A.: Using engine signature to detect metamorphic malware. In: Proceedings of 4th ACM workshop on Recurring malcode (2006).
14. Grosser, Tobias, et al.: Polly-Polyhedral optimization in LLVM. In: Proceedings of the First International Workshop on Polyhedral Compilation Techniques, IMPACT (2011).
15. Rutkowska, Joanna.: Introducing stealth malware taxonomy. COSEINC Advanced Malware Labs, 1–9 (2006).

Object-Based Graphical User Authentication Scheme

Swaleha Saeed and M. Sarosh Umar

Abstract The technique of user authentication remains a key issue over the decades. The main motive behind proposal of graphical password is the human inclination to remember images better than text. In this paper, we have proposed a graphical user authentication scheme that is a hybrid technique, combination of recognition-based scheme and dynamic graphics consisting of objects. The objectives of the proposed technique are to resist shoulder surfing attacks, guessing attacks, etc., without compromising the usability. User study shows that the proposed technique is robust, secure, also offers high usability, and memorability. The results demonstrated that the scheme do not require any additional hardware and can be easily implemented in existing set-up, hence suited for authentication in public places such as ATMs, cyber cafes, mobile phones, etc.

Keywords Cognometric schemes · Dynamic graphics · Graphical user authentication · Shoulder surfing attack

1 Introduction

Security is one of the most relevant areas of concern in today's world of network technology. As a result, authentication remains the key issue in most computer security contexts. Hence, secure authentication becomes the basic necessity for human computer interaction applications. The knowledge based authentication techniques are the most widely used authentication approaches, which can be further decomposed into text and graphical passwords. Text passwords have gained the importance over the years because of its inexpensive implementation and ease

Swaleha Saeed (✉) · M. Sarosh Umar
Department of Computer Engineering, Aligarh Muslim University,
Aligarh, India
e-mail: swalehasaeed@zhcet.ac.in

M. Sarosh Umar
e-mail: saroshumar@zhcet.ac.in

of use [1, 2]. However, research analyses proved that text-based approaches need to satisfy two simultaneous conflicting requirements, i.e., easy to use and hard to guess [1–4]. Ease of remembrance motivates the users to choose easily guessable short passwords leading to the threat of security attacks such as brute-force and dictionary attacks. Reinforcement of strong password promotes them to choose some difficult passwords which could be difficult to remember. This makes the user to write his/her password on sticky notes exposing them to direct theft [1, 4]. These limitations of textual passwords lead to the progression of graphical password as a promising alternative to text-based approaches. The idea of graphical password was proposed by Greg Blonder in 1996 [2, 5]. Till now, various graphical password schemes have been proposed so far. Psychology studies proved that the human brain can easily recognize and recall images better than text for longer period of time [6–9]. Thus, graphical passwords are intended to reduce the memorability and security issues.

In this paper, we propose a new hybrid technique—recognition-based scheme combined with dynamic graphics (having color balls as objects). The preliminary security analysis was carried out which demonstrates that the proposed scheme is robust, memorable, and resistant to shoulder surfing attack.

The paper is organized as follows. Section 2 describes the functionality of proposed scheme. Section 3 deals with security aspect while Sect. 4 includes conclusion of the work.

2 Proposed Scheme

The proposed scheme is a hybrid approach—combination of dynamic graphics and recognition based techniques. Recognition schemes, also known as cognometric schemes, basically deal with identifying user images from an image portfolio. Here, dynamic graphic comprises of different color balls which keeps on changing the color every second. During authentication, user needs to recognize the password images from set of decoy (or distracter) images. The main objective of this approach is to enhance the usability without compromising the security. Basically, the technique comprises two phases: registration phase and authentication phase as described in this section.

2.1 Registration Phase

In registration phase, user creates his/her account by entering user name and then he/she is prompted to select password images (maximum 5 images) from the presented 16 images, arranged in 4×4 grid as shown in Fig. 1. Below every image a 3-digit random code is displayed, so user has to enter the associated code in the text box in order to select that image. The user could have clicked onto images directly,



Fig. 1 Registration phase

however, that results in security breach on account of shoulder surfing attacks. Entering some codes in the text box easily deludes prying eyes [10–13]. User must remember the order of selection of image for his/her password. The random code is displayed only in the registration phase. During later authentication phase, the images are associated with color balls as described in the subsequent section.

2.2 Authentication Phase

The authentication phase in this scheme is different from any other existing graphical password scheme proposed so far. Authentication or login phase deals with identifying the right user. In the proposed scheme, this phase is further divided into two subphases: login phase I and login phase II.

In login phase I, 16 images of the same image set are again presented in random position and below each image an object (ball of a particular color in the proposed scheme) is depicted. This random assignment of color balls to image portfolio is session independent, so the user has to remember image-ball combination only for

that session which enhances the security of the technique. Here, in this scheme there are five color balls (red, green, yellow, blue, and black), which is randomly assigned to the 16 images challenge set. This assignment of image-ball combination in login phase reduces the memorability issues as the user has to remember the image-ball combination only for that session. However, if we had incorporated this idea in the registration phase then the user must remember the password image and also their associated color balls. The first stage is depicted in Fig. 2.

After clicking OK button, the login phase II begins. In login phase II, again the same image set is presented in 16×1 grid form. Here, the color of the balls associated with the images keep changing every second. This phase is a multilevel phase and is presented five times. In level 1, the user has to recognize his/her first image of password set and as soon as the corresponding color ball appears, he/she has to hit next button within a specific time frame. Now the scheme proceeds to the next level, i.e., level 2 wherein, 16×1 image grid is again repeated and the user has to remember the color of the ball associated with his/her second image as shown in Fig. 3. Subsequently phase II commences and the user has to click at the instant the ball acquires the same color as displayed in phase I. This process is repeated five times irrespective of the fact the user clicks at the right color instant and in correct order which marks the end of login phase.

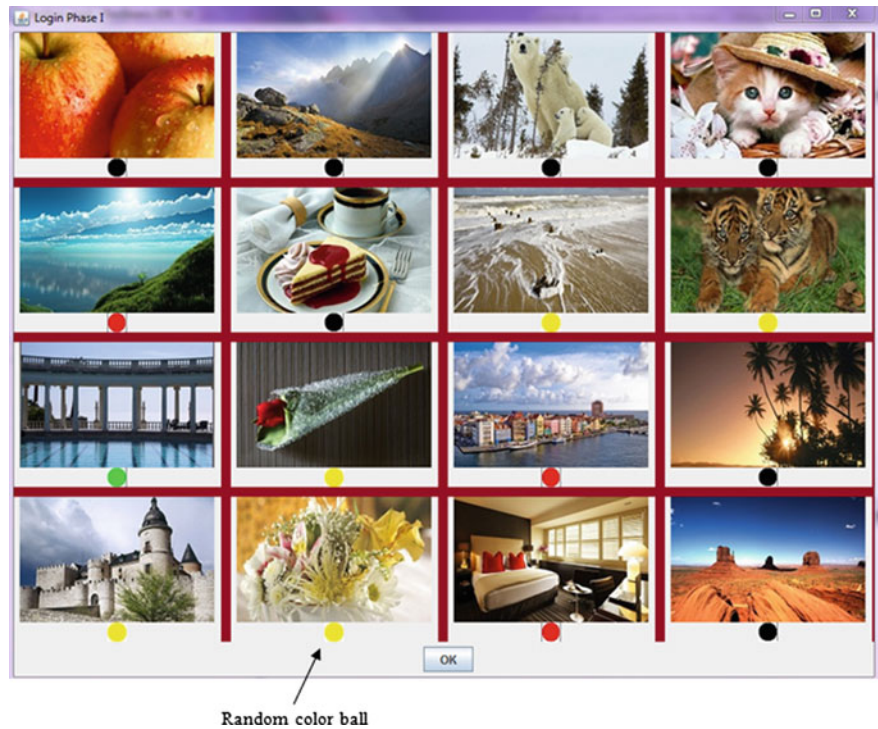


Fig. 2 Login phase I



Fig. 3 Login phase II—level 1

To make the scheme more resistive to security threats like brute-force attack, shoulder surfing attack, dictionary attacks, etc., time factor is also incorporated in the scheme. Thus, a time slot of 15 s is associated with each level of login phase II. The color of ball changes per second, after 5 s the same color pattern is repeated. Hence, user has three trials to correctly identify the color of the ball in 15 s of time frame for that level. The user has to recognize his/her image and associated color within these trials. However, if he/she is unable to complete this identification process within the time slot, a *Time-up* message will appear which ends the authentication process. If the user is able to successfully identify his/her password image set along with correct color ball then he/she is authenticated.

The decomposition of login phase II into five levels make the authentication process uniform for all users and at the same time difficult for attacker to guess the password length.

3 Security Analysis

We have conducted user study on heterogeneous population to investigate the performance of proposed scheme. There were 48 participants/users in the investigation process, 28 were university students and 20 were general computer users comprising nonengineering students, employees, and children. First, a proper training session was provided to them in which they were allowed to interact with scheme by creating their own accounts. The participants practiced authentication phase few times likely 9 or 10. After then, three interaction sessions were designed to explore the feasibility and security aspects of proposed scheme. The first interaction session started after the training session on the same day, i.e., day 0. The second interaction session took place on day 8 and the last session was conducted on day 20. The responses obtained by participants throughout the session were accumulated based on that analysis was performed.

3.1 Performance Analysis

The performance of the system can be measured in two aspects, ease of use (lesser login time) and memorability. The login time can be defined as the time duration when server receives login request to the time when server gave its response.

The login time for each participant was recorded on three respective interaction sessions—day 0, day 8, and day 20. The login time for 10 users obtained is shown in Fig. 4 (because of limited space we have shown only for 10 users, however the analysis was carried out for 48 participants). It can be concluded from the graph (for 10 user) that the average login time recorded on day 0 is 6.2 s, for day 8 it is 6.6 s, and on last session, i.e., day 20 it rises to 7.2 s. The login time of proposed work is compared with other shoulder surfing resistant scheme; for mouse clicking technique [1] it is 31.13 s, whereas for SSP scheme [3] the login time is 41.13 s. The smaller login time obtained proved that they were able to quickly interact with the system irrespective of their background knowledge.

3.2 Guessing Attack Analysis

Guessing attack can be defined as applying trial and error method for identifying user’s password. This attack exhaustively guesses all passwords within the password space (number of options available to users for choosing password in the scheme). To minimize the threat of guessing attack the size of password space should be large enough. For N character alphanumeric password the password space comes out to be 94^N [5, 14, 15]. For a recognition-based graphical authentication of 100 images set out of which N has to be selected the password space is given by $C(100,N)$. From Table 1, it is clear that the proposed scheme offers better resistance to guessing

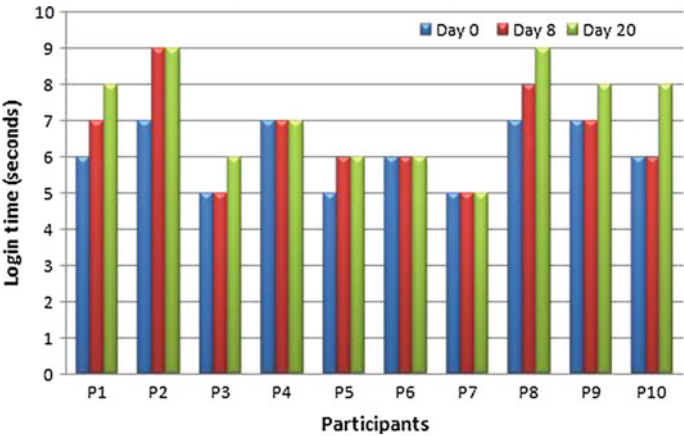


Fig. 4 Login time graph

Table 1 Password space comparison

Technique name	$N = 3$	$N = 4$	$N = 5$
Text-based schemes	2^{20}	2^{26}	2^{33}
Graphical schemes	2^{20}	2^{22}	2^{26}
Proposed scheme	2^{75}	2^{100}	2^{125}

attacks as compared to other approaches. The password space of proposed scheme can be calculated based on (1).

$$\sum_{N=1}^5 (n \times c \times t \times b)^N$$

(1)

- where
- n is total number of images displayed in phase 1
 - c is the total number of distinct colored objects
 - t is the time duration (ms) of each window in phase 2
 - b is number of images displayed in phase 2
 - N is the number of images in user password.

3.3 Shoulder Surfing Attack Analysis

In shoulder surfing attack, the attacker may gain the knowledge of user’s credentials through direct observations or indirectly recording their interaction with external device such as video cameras [5]. Generally, the graphical password schemes utilize keyboard and mouse as their password input devices that are easy to be observed by a shoulder surfing attacker. Also, high resolution cameras with telephoto lenses and equipments make shoulder surfing attack a great concern.

To analyze shoulder surfing attack we have considered the following scenario. The attacker can capture screen shots of five successful authentication rounds across various sessions and can then compare them to obtain user’s password images. However, he or she will find that the time instant at which the authorized user clicks varies in all the sessions. Thus it is difficult for him/her to know the exact instant at which the click operation is to be performed as the color of the ball associated with each image changes every second and the simple screen shots will not suffice to determine the correct time tapping instant. Also, the color assignment to balls is totally random and it is not necessary that the same color appears at the same time in next login session. This random fusion of image portfolio and objects (color balls) makes the proposed scheme highly repellent to shoulder surfing attack. The success probability of shoulder surfing attack is calculated for proposed scheme for three interaction sessions and compared to the other schemes as summarized in Table 2. The values obtained from Table 2 reflect that the proposed work offers great resistance to shoulder surfing attack compared to the other two techniques.

Table 2 Shoulder surfing attack success probability

Technique name	Success probability
Mouse clicking [1]	1.28
SSP scheme [3]	0.56
Proposed scheme	0.02

4 Conclusion

This paper introduces a novel object-based graphical password authentication technique which exploits the idea of recognition-based approach coupled with the concept of dynamic graphics in time domain. Because of the assimilation of time factor in authentication phase the scheme effectively resist shoulder surfing attack. Here, authentication relies not merely on click operation but also on the click-time instant when the desired color ball (object) appears on the screen. The results obtained through user study prove that the scheme is easy to use and resistive to security attacks. Further, the analysis carried out demonstrated that the scheme offers larger password space to withstand against human guessing attack. Based on these evaluation strategies, it can be concluded that the technique is able to resolve security-usability conflict issue efficiently. The idea can be easily adopted in different authentication application like, ATMs, mobile phones, access control without any changes in existing set-up.

References

1. Wiedenbeck, S., Waters, J., Sobrado, L., Birget, J.C.: Design and evaluation of a shoulder-surfing resistant graphical password scheme. In: Proc. of the Working Conference on Advanced Visual Interfaces, pp. 177–184 (2006).

2. Chiasson, S., Forget, A., Biddle, R.: Accessibility and graphical passwords. In: Symposium on Accessible Privacy and Security, Pittsburgh, USA (2008).

3. Wu, T.S., Lee M.L., Lin, H.Y., Wang, C.Y.: Shoulder-surfing-proof graphical password authentication scheme. In: International Journal of Information Security, vol. 13, issue 3, pp. 245–254. Springer Berlin, Heidelberg (2014).

4. Alsaleh, M., Mannan, M., Oorschot, P.C.V.: Revisiting defenses against large-scale online password guessing attacks. In: IEEE Transaction on Dependable and Secure Computing, vol. 9, no. 1, pp. 128–141 (2012).

5. Oorschot, P.C.V., Salehi-Abari, A., Thorpe, J.: Purely Automated Attacks on PassPoints-Style Graphical Passwords. In: IEEE Transaction on Information Forensics and Security, vol. 5, no. 3, pp. 303–404 (2010).

6. Umar, M.S., Rafiq, M.Q.: A Graphical Interface for User Authentication on Mobile Phones. In: ACHI 2011: The Fourth International Conference on Advances in Computer-Human Interactions, pp. 69–74, Guadeloupe, France (2011).

7. Eluard, M., Maetz, Y., Alessio, S.: Action-based graphical password: Click-a-Secret. In: 2011 IEEE International Conference on Consumer Electronics, pp. 265–266 (2011).

8. Sabzevar and Stavrou: Universal multi-factor authentication using graphical passwords. In: Proc. of the 2008 IEEE International Conference on Signal Image Technology and Internet Based Systems, pp. 625–632 (2008).

9. Umar, M.S., Saeed, S.: Graphical User Authentication Based on Image Transformation. In: Proc. of 3rd International Conference of Engineering and Applied Sciences, pp. 7–14 Alberta, Canada (2014).
10. Lin, R., Huang, S.Y., Bell, G.B., Lee, Y.K.: A new CAPTCHA interface design for mobile devices. In: Proc. 12th Austral. User Inter. Conf, pp. 3–8 (2011).
11. Syed, Z., Banerjee, S., Cheng, Q., Cukic B.: Effects of user habituation in keystroke dynamics on password security policy. In: Proc. IEEE 13th Int. Symp. High-Assur. Syst. Eng., pp. 352–359 (2011).
12. Bonneau, J.: The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In: Proc. IEEE Symp. Security Privacy, pp. 20–25 (2012).
13. Trewin, S., Swart, C., Koved, L., Martino, J., Singh, K., Ben-David, S.: Biometric authentication on a mobile device: A study of user effort, error and task disruption. In: Proc. 28th Annu. Comput. Security Appl. Conf., pp. 159–168 (2012).
14. Alsulaiman, F.A., Saddik, A.E.: Three-Dimensional Password for More Secure Authentication. In: IEEE Transaction on Instrumentation and Measurement, vol. 57, no. 9, pp. 1929–1938 (2008).
15. Shi, Zhu, Youssef: A PIN entry scheme resistant to recording-based shoulder-surfing. In: Proc. of the 2009 Third International Conference on Emerging Security Information, Systems and Technologies, pp. 237–241 (2009).

Efficient Density-Based Clustering Using Automatic Parameter Detection

Priyanka Sharma and Yogesh Rath

Abstract Clustering governs huge data by organizing similar data objects into groups. Density-based clustering permits composition of data objects on basis of their density distribution. DBSCAN, an illustrious and prominent density-based clustering algorithm gives birth to arbitrary-framed clusters, without requiring preexisting acquaintances on the number of clusters to be produced. The inputs to DBSCAN principal are: dataset required to be mined, radius of neighborhood—*Eps* (ϵ), minimum number of points needed to build a cluster (*MinPts*). *DBSCAN* clustering desires these two parameters to be given as input manually and automatic detection of these parameters is a very tedious exercise and has a significant influence on clustering result. In this paper, we contemplated a new and efficient density-based clustering algorithm (*E-DBSCAN*). The consolidated notion of the proposed approach is that it avoids manual intervention of input values. Experimental results demonstrate effectiveness and efficiency of the proposed algorithm on varied domain of datasets.

Keywords Cluster analysis • DBSCAN • Eps • MinPts • Parameter detection

1 Introduction

Diverse areas desire the administration of data in such a way that fast growing amount of data can be managed successfully. It derives a need for some methods to extract knowledge from data. In the literature, various assignments of knowledge discovery in databases (*KDD*) [1] have been illustrated. Cluster analysis accommodates a set of abstract/physical objects into classes of equivalent objects. Clusters

Priyanka Sharma (✉) • Yogesh Rath
Department of Computer Science & Engineering, Yagyavalkya Institute of Technology,
Jaipur, India
e-mail: priyanka08291@gmail.com

Yogesh Rath
e-mail: yogeshrathicsyt@gmail.com

harmonize to obscure patterns, clustering is a form of unsupervised learning of a mysterious data concept which intends to find resemblances in training data which is used for initiatory and fundamental data analysis for finding the hidden patterns or grouping the data into chunks. These divisions (clusters) are composed by using an allowance of similarity. In density-based methods, clusters are regarded as dense regions scattered by noise (regions of low density). These methods are capable of finding clusters, noise, outliers, etc. Some leading density-based approaches are *DBSCAN*, *DBCLASD*, *OPTICS*, *DENCLUE*, *MAFIA*, and many more. Density-based spatial clustering of applications with noise (DBSCAN) gamble on density-based cluster aspects and originates discretionary sketched clusters and used in many areas such as to determine the reputation of any given application or organization by resolving the zonal followers for narrow zones [2]. DBSCAN requires two input parameters and reinforces the user in determining an applicable value for it. DBSCAN provides some consequential assistance such as:

- Do not require any prior knowledge on the total number of predefined clusters to be formed.
- Arbitrary-shaped clusters are determined, as cluster's shapes in spatial databases may be orbicular, lengthy, continuous, stretched, etc.
- Better efficiency on large databases, i.e., on databases having more than thousands of objects.
- Treatment of noisy data and outliers.

DBSCAN concludes into proper clusters but there are two distinct deficiencies for it:

- The clustering realization depends on two specified parameters—the radius of a neighborhood and the minimum number of the data points contained in such neighborhood. These parameters serve a distinguished density. Without enough precedent knowledge, these two parameters are difficult to determine.
- Using these parameters for a single density, DBSCAN does not deliver goods to datasets with varying densities.

The intent is to adduce a method which uses a technique to automatically detect these two parameters. Using basic density-based clustering algorithm (DBSCAN) we can not automatically detect these input values, also some most trendy algorithms like *k*-means desire the input value *k* (number of clusters formed in clustering process) that has to be given as input by user.

We are using varied density datasets as input. The values *Eps* neighborhood and *MinPts* (minimum number of points in a cluster) will be determined by the algorithm. The output will show clustering of the input dataset. The dataset can contain two or multidimensional attributes. Here, we use various datasets for the algorithm to execute. We will compare the output with the existing DBSCAN and put the improved DBSCAN for the above-mentioned problem. The proposed one will have better output and reduced time complexity.

The paper is formulated as follows: Sect. 2 outlines the literature survey or related works. Section 3 illustrates the proposed approach. Section 4 depicts experimental results. Finally, Sect. 5 defines the conclusion.

2 Related Work

DBSCAN [3] selects any random point p (not visited) and extracts its neighbors using neighborhood radius— Eps . If sufficient points exist around p , it yields a cluster. If not, DBSCAN move to next point and the procedure is repeated until all points are traversed. Points in DBSCAN are classified as core point and border point. If a point contains at least $MinPts$ in its Eps neighborhood then it is core point otherwise it is noise.

Mohammed et al. [4] proposed DMDBSCAN, in which different shaped and sized clusters are determined that may differ on the basis of their local densities. It chooses multiple values of Eps neighborhood for various densities depending on a k -dist plot. Now DBSCAN algorithm is executed for each value of Eps to ensure that all the points are clustered.

In [5], authors proposed a research on adaptive parameter determination. The idea is that the values of parameters Eps and $MinPts$ are ascertained based on the statistical properties of the dataset. A distance distribution matrix $DIST_{n \times n}$ is calculated, where n is the number of objects in the dataset D . $DIST_{n \times n}$ is a real symmetric matrix with n rows and n columns, in which each element denotes the distance between objects i and j in D .

Glory [6] extends the research by focusing on another major flaw of DBSCAN that it is unable to determine clusters that survive within another cluster. Initially, Eps , and $MinPts$ are set on lower values and then raised bit by bit in next steps. Euclidian and Manhattan distances are used by the author.

In [7], authors proposed an enhanced DBSCAN algorithm. To determine different range of Eps values automatically authors addressed a k -dist graph for all the points. The average of the distances of every point to all k of its nearest neighbors is computed initially. To determine $MinPts$, authors calculated the number of data objects in Eps neighborhood of every point in dataset one by one. And then mathematic expectation of all these data objects is calculated, which is the value of $MinPts$.

Kedar [8] evolved a new method which determines the value of Eps using the value of ' k ' in varied density-based spatial cluster analysis. A k -dist graph is drawn for all points and the average of the distances of a point to all k of its neighbors is computed. These average k -distances are plotted in ascending order. A knee (threshold) is determined when a sharp change occurs in the plot. The algorithm makes use of average determination and distance measurement.

DBSCAN is also generalized in the way authors [9] have used. The approach detects the input parameter epsilon, which enables DBSCAN to get unconstrained

of any attainment. It relies on the data distribution of each dimension; due to this relevant feature the approach is also applicable for subspace clustering.

Amin et al. [10] presents an efficient and effective hybrid clustering method *BDE-DBSCAN* that combines binary differential evolution and DBSCAN algorithm to simultaneously, quickly, and automatically specify appropriate parameter values for *Eps* and *MinPts*. Since the *Eps* parameter can largely degrade the efficiency of the DBSCAN algorithm, the combination of an analytical way for estimating *Eps* and tournament selection (TS) methods is also employed.

Jamshid et al. [11] proposed an enhancement which will remove the radius—*Eps* and replace it with some another parameter p (noise ratio of the dataset). The author described that this method will not reduce the number of parameters but the p parameter is usually much simpler to set than *Eps*, because in some applications, the user knows the noise ratio of the dataset in advance.

3 Proposed Method

Nowadays, the volume of data is growing exponentially. Clustering takes care of data to some extent by grouping similar data objects. Still, the dilemma that arises with DBSCAN is that it is unable to detect parameters implicitly and it is very susceptible to clustering parameters. Also it breaks down to deal with varied densities and too sparse dataset. Here, the proposed algorithm *EDBSCAN* (efficient density-based spatial clustering of applications with noise) will find out the parameters for efficient clustering results. The dominant tasks of *EDBSCAN* are:

1. The object which to be clustered is subdivided into different multiple cells, *Eps* and *MinPts* are calculated, respectively for each cell. Now, two scenarios can be there; whether or not the cell contains any input objects.
2. Finally, the *Eps* and *MinPts* pairs are merged. Based on these values, now the objects will be clustered (Merging).

In Fig. 1a, the objects are partitioned into a total of 16 rectangular cells. The cells are numbered from 1 to 16. As we can see, the cells 1, 4, 5, 7, 13, 14, 15 contain the maximum proportion of objects. Cells 2, 10, 16 contain less objects and remaining cells 3, 6, 8, 9, 11, 12 contain the least amount. Objects in cell 1, 4, 5, 7, 13, 14, 15 are closest to each other. Green color objects, i.e., with less density may consider being noise by DBSCAN. Using only a single pair of *Eps* and *MinPts*, DBSCAN may detect three clusters as shown in Fig. 1a.

The idea of *EDBSCAN* is to generate *Eps* and *MinPts* pairs for each cell. So, a total of 16 pairs of *Eps* and *MinPts* will be generated (Fig. 1b). *Eps* and *MinPts* pairs of cell 2, 10, 16 and 1, 4, 5, 7, 13, 14, 15 and 3, 6, 8, 9, 11, 12 are merged (according to densities). Now, these three pairs are used to produce clusters. *EDBSCAN* determines eight clusters altogether.

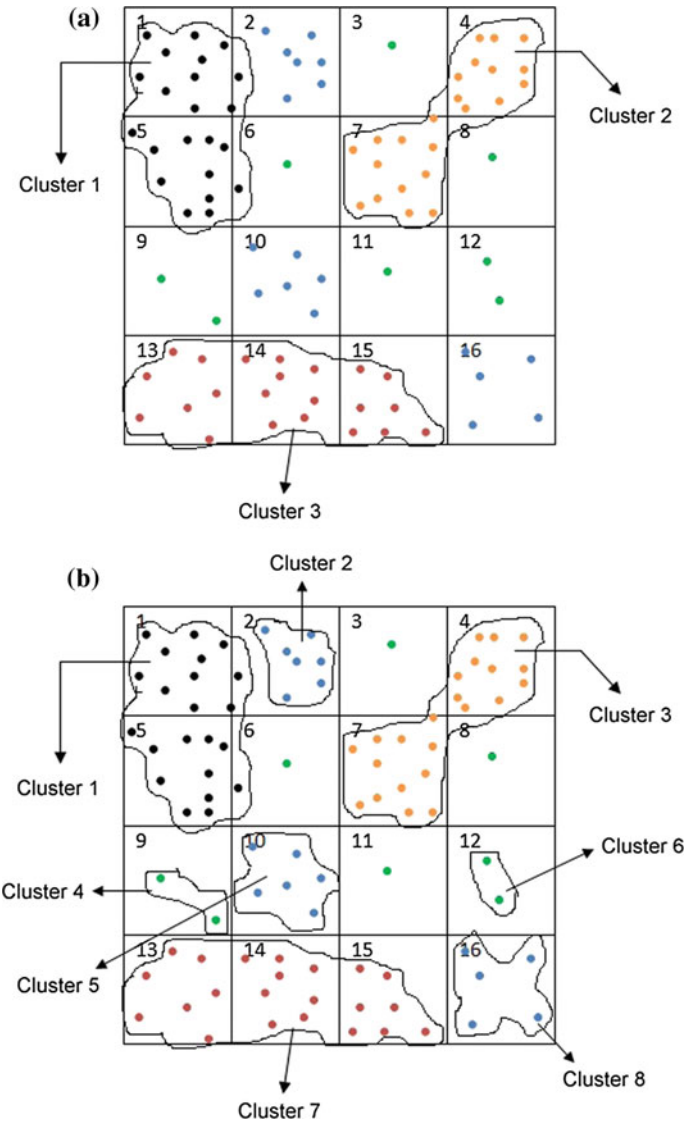


Fig. 1 **a** Single pair of *Eps* and *MinPts* with clusters having similar densities. **b** Multiple pairs of *Eps* and *MinPts* with clusters having different densities (*EDBSCAN*)

3.1 Automatic Parameter Generation Using *EDBSCAN*

The proposed algorithm is the advancement over *DBSCAN* as it will automatically find clustering parameters. As compared to original *DBSCAN* method, *EDBSCAN* algorithm will work efficiently and in a well-behaved manner. The modified

algorithm is demonstrated here in which DS denotes the dataset. The neighborhood radius is specified by Eps or ε and minimum numbers of points are shown as $MinPts$.

Algorithm: The pseudo code of proposed technique (<i>EDBSCAN</i>) to find suitable ε and $MinPts$ for each level of density in data set.	
Purpose:	To find suitable values of ε and $MinPts$
Input:	DS - Data Set of size n
Output:	Eps and $MinPts$ for each varied density
Procedure:	EDBSCAN ($D, Eps, MinPts$) 1. Create a k-d tree. 2. Calculate the distance $DIST$ between each pair of objects. 3. Split objects into a number of cells n. The Cell Set is defined as : $CS = \{C_k \mid k = 1, 2, \dots, NOC\},$ $C_k = N_O_k$ Where NOC is the number of cells and N_O_k is the number of objects in the k^{th} cell. 4. If the cell does not contain any input object (i.e. $IO_k = \emptyset$), then calculate Eps_k and $MinPts_k$ as: $\varepsilon_k = \frac{\sum_{kcc=1 \dots N_O_k} \varepsilon_{kc}}{ N_O_k \beta}, \quad \varepsilon_{kc} = \min (d(O_j, O_i)) (i \neq j), \quad IO_k = \phi$ $MinPts_k = \text{median } EN(O_i, \varepsilon_i) , \quad IO_k = \phi$ Where $EN(O_i, \varepsilon_i)$ is the ε_i neighborhood of O_i. 5. Else if, m objects are submitted by user (i.e. $IO_k \neq \emptyset$), calculate Eps_k and $MinPts_k$ as: $\varepsilon_k = \frac{\sum_{keO_e \in IO_k} \varepsilon_{ke}}{ IO_k \beta}, \quad \varepsilon_{ke} = \min (d(O_j, O_e)) (e \neq j), \quad IO_k \neq \phi$ $MinPts_k = \frac{\sum_{O_e \in IO_k} EN(O_i, \varepsilon_i) }{ IO_k }, \quad IO_i \neq \phi$ 6. Merge same value pairs in order to reduce complexity. 7. Update Eps_k and $MinPts_k$. 8. for $i = 1$ to NEM { Apply DBSCAN

4 Experiments and Results

Modern era constitutes substantial research findings in the field of clustering on varied domain of datasets. In the experiments, we have tested the proposed method using two datasets (Sample Dataset and Twitter Follower's Dataset) and found the best clustering of objects and compared the results with the original DBSCAN algorithm. The implementation is performed on VB.net using VS Ultimate2013 and includes both the manual insertion of parameters as well as their automatic detection. All experiments are run on a system having 2.30 GHz processor and 3 GB RAM.

4.1 Comparisons

4.1.1 Input Independency

DBSCAN: insists users to give *Eps* and *MinPts*, which is tedious job for users.

EDBSCAN: EDBSCAN is entirely independent on input, as employs a way to automatically detect parameters.

4.1.2 Noises Detection

DBSCAN: the core DBSCAN procedure is unable to detect noises properly for a large quantity of objects. The ratio of noise detection depends on the inputs provided.

EDBSCAN: proposed algorithm EDBSCAN is able to properly detect noises, as it does not rely on input parameters.

4.1.3 Clusters Having Varying Densities

DBSCAN: produce single pair of *Eps* and *MinPts* clusters having similar density, as demonstrated in Fig. 1a.

EDBSCAN: originates multiple pairs of *Eps* and *MinPts* having varied densities, as demonstrated in Fig. 1b (Fig. 2 and Table 1).

4.2 Results

Here the input data parameters are detected automatically and multiple pair of these parameters having varying densities are originated. Figure 2 and Table 1 shows the result, when applied on above mentioned datasets.

Fig. 2 Comparative analysis of clustering algorithms

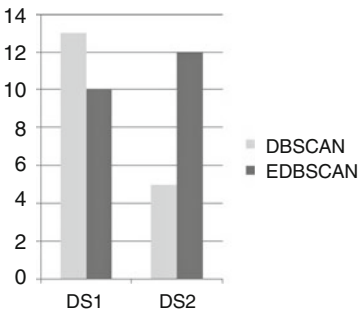


Table 1 Comparison of clustering algorithms

Algorithms	Input	No. of clusters	Objects in a cluster (%)
DBSCAN	{0.03,200}	13	5, 5, 4, 9, 4, 5, 4, 7, 3, 4, 3, 6, 6
	{0.05,500}	5	5, 10, 8, 7, 7
EDBSCAN	ϕ	10	11, 14, 13, 4, 5, 4, 7, 7, 3, 6, 6
	ϕ	12	9, 6, 14, 10, 4, 5, 5, 8, 4, 3, 7, 7

5 Conclusion and Future Work

The eminent algorithm DBSCAN goes through the flaw: manual intervention of input parameters. This paper proposed a method for automatic origination of density-based parameters. The proposed algorithm is efficient for varied domain of datasets in terms of performance and complexity. We aimed at improving the efficiency of original density-based clustering algorithm. The results are presented for the experiments performed for examining various datasets against original DBSCAN and the proposed DBSCAN. The work done can also be enhanced in future to get more accurate results also for high dimensional datasets.

References

1. Tan, Michael Steibach, Vipin Kumar, “Introduction to Data Mining”, Third impression by Dorling Kidersley, 2009.

2. N. Jain, B. Agarwal, M. K. Gupta, “Improved DB-SCAN for Detecting zonal Followers for Small Regions on Twitter”, International Conference on Information systems Design and Intelligent Applications, Springer, 2015.

3. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” Proc. Second Int’l Conf. Knowledge Discovery and Data Mining (SIGKDD), 1996.

4. Mohammed T. H. Elbatta and Wesam M. Ashour, “A Dynamic Method for Discovering Density Varied Clusters”, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.-6 (1), 2013.

5. Hongfang Zhou, Peng Wang, Hongyan Li, "Research on Adaptive Parameters Determination in DBSCAN Algorithm", *Journal of Information & Computational Science* Vol.-9, pp. 1967–1973, 2012.
6. Glory H. Shah, "An improved DBSCAN, a density based clustering algorithm with parameter selection for high dimensional data sets", *International Conference on Engineering (NUICONE)*, IEEE, Vol-7, pp. 1–6, 2012.
7. Manisha Naik Gaonkar, Kedar Sawant, "AutoEpsDBSCAN:DBSCAN with Eps Automatic for Large Dataset", *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, Vol.-2, pp. 2319–2526, 2013.
8. Kedar Sawant, "Adaptive Methods for Determining DBSCAN Parameters", *International Journal of Innovative Science, Engineering & Technology*, Vol.-1(4), 2014.
9. S. Jahirabadkar, P. Kulkarni, "Algorithm to determine e-distance parameter in density based clustering", *Expert Systems with Applications*, Vol.-41, pp. 2939–2946, 2014.
10. Amin Karami, Ronnie Johansson, "Choosing DBSCAN Parameters Automatically using Differential Evolution", *International Journal of Computer Applications*, Vol.-91 (7), 2014.
11. Jamshid Esmaelnejad, Jafar Habibi, Soheil Hassas Yeganeh, "A Novel Method to find Appropriate Eps for DBSCAN", *Proceedings of the Second international conference on Intelligent information and database systems*, Springer, pp. 93–102, 2010.

WT-Based Distributed Generation Location Minimizing Transmission Loss Using Mixed Integer Nonlinear Programming in Deregulated Electricity Market

Manish Kumar, Ashwani Kumar and K.S. Sandhu

Abstract In this paper, analysis has been carried out for transmission loss minimization with the integration of wind turbine in the power system network. For obtaining the power output from the wind turbine, a probabilistic model for the wind has been considered using Weibull distribution function. A mixed integer nonlinear programming (MINLP) approach has been utilized for determining optimal location and number of distributed generators considering minimization of transmission loss. The main objective of the paper is: (i) transmission loss minimization with wind turbines, (ii) optimal location and sizing of wind turbines, (iii) comparison of results without and with wind turbines considering constant load model and realistic ZIP load model. The analysis has been carried out with constant P, Q load and the realistic ZIP load model. The impact of different load models has also been studied. The proposed MINLP-based optimization approach has been applied for IEEE 24 bus reliability test system.

Keywords Distributed generation · Mixed integer nonlinear programming · Optimal location · Wind turbine

Manish Kumar · Ashwani Kumar (✉) · K.S. Sandhu
National Institute of Technology, Kurukshetra 136119, India
e-mail: ashwani.k.sharma@nitkkr.ac.in

Manish Kumar
e-mail: khanagwal.manish@gmail.com

K.S. Sandhu
e-mail: kjssandhu@rediffmail.com

1 Introduction

The distributed generation has lot of advantages as it may help to reduce losses in the network due to its availability near to the load centers. It will increase reliability of the network and improves the voltage profile in the distribution and the transmission systems [1]. Many authors presented renewable energy integration issues with the aim of power loss reduction using optimization based on genetic algorithm (GA) [2–5]. In [6] two stage optimization-based approach for distributed generation was presented. Authors in [7] proposed mixed integer programming based approach for DG placement. Optimal planning of DGs in the distribution network was proposed in [8]. Many other approaches based on the heuristic approaches and optimal energy mix for loss minimization in [9, 10]. Probabilistic approach for wind energy allocation was proposed in [11]. An analytical expression for distributed generation allocation for loss minimization was proposed in [12]. Multiple distributed generation placement was proposed in primary distribution systems [13, 14]. Wind speed prediction, modeling, and forecasting were proposed in [15, 16].

In this paper, wind-based distributed generator (DG) is used for connecting the power system and to reduce the transmission losses. A mixed integer nonlinear programming (MINLP) approach for deterring optimal location and number of distributed generators considering minimization of transmission loss has been proposed. The total real and reactive power loss, percentage reduction in active power loss, and optimal DG location has been obtained. The results have also been obtained for minimization of transmission losses with ZIP load, also considering different cases of ZIP load coefficients namely same ZIP load coefficient and different coefficients at each load bus. The optimization approach has been applied for IEEE 24 bus reliability test system.

2 Wind Turbine Generation Pattern Modeling

The variations of wind speed v are modeled as a Weibull PDF and its characteristic function which relates the wind speed and the output of a WT is as follows:

$$\text{PDF}(v) = \left(\frac{k}{c}\right) \left(\frac{v}{c}\right)^{(k-1)} \exp\left[-\left(\frac{v}{c}\right)^k\right] \quad (1)$$

$$k = \left(\frac{\bar{\sigma}}{\bar{\mu}}\right)^{-1.086} \quad (2)$$

$$c = \frac{\bar{\mu}}{\Gamma\left(\frac{1}{k} + 1\right)} \quad (3)$$

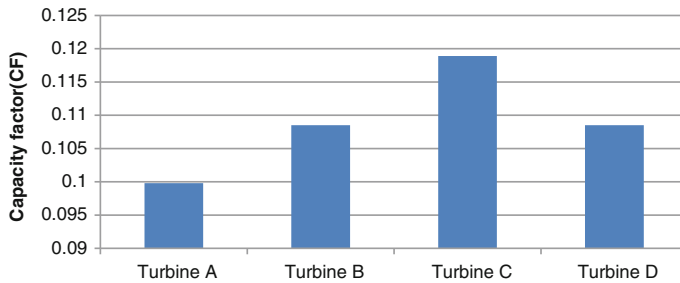


Fig. 1 Capacity factor (CF) of wind turbine available

where k and c are the shape and scale factor of the Weibull PDF of wind speed $\bar{\sigma}$ and $\bar{\mu}$ is mean m/s and standard deviation m/s, we used the data for the hourly mean with speed during the month of May over the first 12 years (1994–2005) in [17]. The hourly wind speed sample has been obtained using the Monte Carlo simulation (MCS). $P_{i,r}^w$: is the rated power of wind turbine installed in bus- i , P_i^w : is the generated power of WT in bus- i , v_{out}^c : is the cut-out speed, v_{in}^c : is the cut-in speed, v_{rsted}^c : is the rated speed of the wind turbine. The speed-power curve of each wind turbine (turbine A, turbine B, turbine C, turbine D) has been obtained. The technical data is given in [10]. The generated power of the wind turbine is determined using its speed-power curve as follows:

$$P_i^w = \begin{cases} 0, & \text{if } v \leq v_{in}^E \text{ or } v \geq v_{out}^c \\ \frac{v - v_{in}^c}{v_{rsted}^c - v_{in}^c} P_{i,r}^w, & \text{if } v_{in}^c \leq v \leq v_{rsted}^c \\ P_{i,r}^w, & \text{else} \end{cases} \quad (4)$$

The value of K (m/s) and C (m/s) hourly are calculated the using Eqs. (2), (3). In this work, we use the average value of active and reactive power generation of each turbine. Figure 1 shows the capacity factor (CF) of each turbine.

$$\text{Capacity factor (CF)} = \frac{\text{Average output power of turbine}}{\text{rated power of turbine}} \quad (5)$$

3 General OPF Formulations in the Presence of Wind Turbine-Based Distributed Generation with Constant P, Q, Load

General objective function for minimization of transmission loss using MINLP approach can be written as:

$$\text{Min } F(x, u, \zeta^{\text{int}}) \quad (6)$$

Subject to equality and inequality constraints is defined as

$$h(x, u, \zeta^{\text{int}}) = 0 \quad (7)$$

$$g(x, u, \zeta^{\text{int}}) \leq 0 \quad (8)$$

where x (state vector of variables V, δ), u (the control parameters, $P_{gi}, Q_{gi}, P_{WTi}, Q_{WTi}$), and ζ^{int} (integer variable with values 0,1). The integer variable zero it represents no DG and one represents presence of distributed generator in the network.

Objective function F is minimization of the transmission loss and can be expressed as sum of the power flows to and from in the transmission lines as:

$$\text{Min } F(x, u, \zeta^{\text{int}}) = P_{ijl} + P_{jil} \quad (9)$$

The line that flows from bus- i to bus- j and bus- j to bus- i is given as

$$P_{ijl} = V_i^2 G_{ij} - V_i V_j (G_{ij} \cos(\delta_i - \delta_j) + B_{ij} \sin(\delta_i - \delta_j)) \quad (10)$$

$$P_{jil} = V_j^2 G_{ij} - V_i V_j (G_{ij} \cos(\delta_i - \delta_j) - B_{ij} \sin(\delta_i - \delta_j)) \quad (11)$$

3.1 Equality Constraints

$$P_i = P_{gi} + \zeta_i^{\text{int}} * P_{WTi} - P_{di} \forall i = 1, 2, \dots, N_b \quad (12)$$

$$Q_i = Q_{gi} + \zeta_i^{\text{int}} * Q_{WTi} - Q_{di} \forall i = 1, 2, \dots, N_b \quad (13)$$

$$P_i = \sum_{j=1}^{N_b} V_i V_j [G_{ij} \cos(\delta_i - \delta_j) + B_{ij} \sin(\delta_i - \delta_j)] \quad \forall i = 1, 2, \dots, N_b \quad (14)$$

$$Q_i = \sum_{j=1}^{N_b} V_i V_j [G_{ij} \sin(\delta_i - \delta_j) - B_{ij} \cos(\delta_i - \delta_j)] \quad \forall i = 1, 2, \dots, N_b \quad (15)$$

3.2 Inequality Constraints

- (a) Real power generation limit of generators at bus- i

$$P_{gi}^{\min} \leq P_{gi} \leq P_{gi}^{\max}, i = 1, 2, \dots, N_g \quad (16)$$

- (b) Reactive power generation limit of generators and other reactive sources at bus- i

$$Q_{gi}^{\min} \leq Q_{gi} \leq Q_{gi}^{\max}, i = 1, 2, \dots, N_q \quad (17)$$

- (c) Voltage limit of $V_i^{\min}, V_i^{\max}$ at bus- i

$$V_i^{\min} \leq V_i \leq V_i^{\max}, i = 1, 2, \dots, N_b \quad (18)$$

- (d) Phase angle limit of $\delta_i^{\min}, \delta_i^{\max}$ at bus- i

$$\delta_i^{\min} \leq \delta_i \leq \delta_i^{\max}, i = 1, 2, \dots, N_b \quad (19)$$

- (e) Line flow limit based on thermal and stability considerations.

$$|S_{ij}| \leq S_{ij}^{\max}$$

3.3 Power Generation Limit of Wind Turbine based Generators at Bus- i

- (a) Real power generation limit

$$P_{WTi}^{\min} \leq P_{WTi} \leq P_{WTi}^{\max}, i = 1, 2, \dots, N_{WT} \quad (20)$$

- (b) Reactive power generation limit

$$Q_{WTi}^{\min} \leq Q_{WTi} \leq Q_{WTi}^{\max}, i = 1, 2, \dots, N_{WT} \quad (21)$$

- (c) Optimal number of distributed generators

$$N_{WT} = \sum_{i=1}^{N_{WT}} \zeta_i^{\text{int}} \leq N_{WT}^{\max} \quad (22)$$

4 ZIP Load Model

The load is modeled as polynomial load as given in [9]:

$$P_{dz} = P_o (A_p V^2 + B_p V + C_p) \quad (23)$$

$$Q_{dz} = Q_o (A_q V^2 + B_q V + C_q) \quad (24)$$

$$(A_p + B_p + C_p) = (A_q + B_q + C_q) = 1 \quad (25)$$

where V is node voltage in p.u, P_o, Q_o the real power and reactive power consumed at the specific node under the reference voltage. A_p, A_q are the parameters for constant impedance (constant Z) load component. B_p, B_q are the parameters for constant current (constant I) load component; C_p, C_q are the parameters for constant power (constant P and Q) load component. The values of A_p, A_q, B_p, B_q and C_p, C_q are determined for different load types in distribution systems.

4.1 Without PV-Based DG

The real and reactive power injection equations can be modified in the presence of ZIP load as

$$P_i = P_{gi} - P_{dzi} \forall i = 1, 2, \dots, N_b \quad (26)$$

$$Q_i = Q_{gi} - Q_{dzi} \forall i = 1, 2, \dots, N_b \quad (27)$$

4.2 With WT-Based DG

With distributed generation, from Eqs. (12) and (13) the real and reactive power constraints are modified in the presence of ZIP load as

$$P_i = P_{gi} + \zeta_i^{\text{int}} * P_{WTi} - P_{dzi} \forall i = 1, 2, \dots, N_b \quad (28)$$

$$Q_i = Q_{gi} + \zeta_i^{\text{int}} * Q_{WTi} - Q_{dzi} \forall i = 1, 2, \dots, N_b \quad (29)$$

5 Results and Discussion

The proposed approach is applied to IEEE 24 bus reliability test system [18] to find optimal distribution generation location. The results have been obtained for voltage profile, total real and reactive power loss, and percentage reduction in the transmission loss with WT-based DGs. The results are also obtained with ZIP load variation at each bus for comparison with constant P, Q load model and it is categorized as:

Case 1 (without WT-based distributed generator), Case 2 (with 1 WT-based distributed generator), Case 3 (with 2 WT-based distributed generators), Case 4 (with 3 WT-based distributed generators), and Case 5 (with 4 WT-based distributed generators).

5.1 Results for Minimization of Transmission Loss with Constant P, Q Load

In Table 1, the results for minimization of constant load are given which contains the total active and reactive loss named as PLT and QLT, respectively. It also

Table 1 Results for minimization of total transmission loss with constant P, Q load

	Case 1: Without DG(WT)	Case 2: with 1 DG (WT)	Case 3: With 2 DG(WT)	Case 4: with 3 DG(WT)	Case 5: with 4 DG(WT)
PLT(p.u.MW)	0.2877	0.2732	0.2609	0.2526	0.2465
QLT(p.u.MVar)	-2.9975	-3.1586	-3.2780	-3.3638	-3.4215
Minimum voltage	0.9661 (At bus 4)	0.9667p.u (At bus 4)	0.9654p.u. (At bus 4)	0.9649p.u. (At bus 4)	0.9689 p.u. (At bus 4)
Total active load (p.u.MW)	28.5	28.5	28.5	28.5	28.5
Total reactive load (p.u.MVar)	5.8	5.8	5.8	5.8	5.8
% age reduction in loss		5.03 %	9.31 %	12.20 %	14.32 %
Optimal Bus l ocation of DG(WT)		3	3,10	3,5,10	3,4,6,10
Total DG(WT) size (p.u.MW)		0.3254	0.6508	0.8679	0.9987
Total DG(WT) size (p.u.MVar)		0.0464	0.0564	0.0664	0.0850

Fig. 2 Voltage profile with and without WT-based DG with constant load

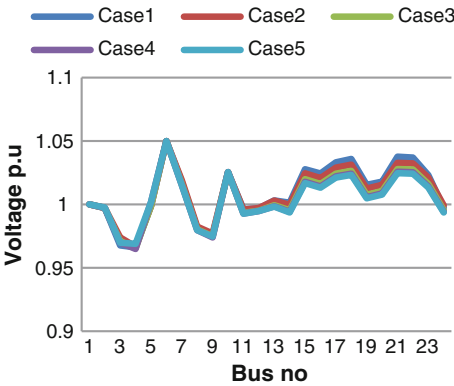


Fig. 3 Total active power loss with and without WT-based DG with constant load

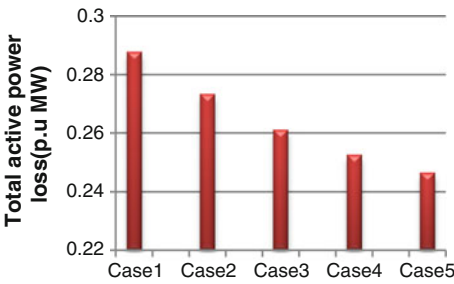
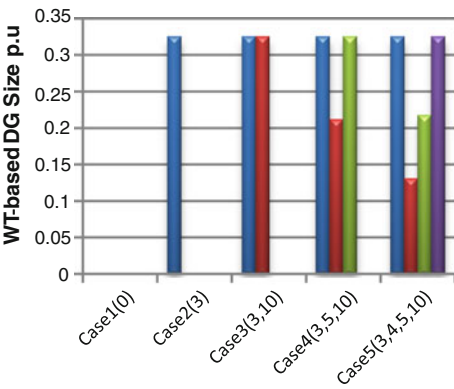


Fig. 4 WT-based DG size in p.u with constant load



represents the percentage reduction in total active power loss which is calculated by the following formula:

$$\% \text{ reduction in loss} = \frac{PLT_{\text{without dg}} - PLT_{\text{with dg}}}{PLT_{\text{without dg}}} \times 100 \% \quad (30)$$

The voltage profile, PLT and WT-based DG size, and location are shown in Figs. 2, 3, and 4, respectively.

5.2 Results with ZIP Load1

The ZIP has been taken at all buses with the ZIP load coefficient as

$$A_p = 0.1, B_p = 0.1, C_p = 0.8, A_q = 0.1, B_q = 0.1, C_q = 0.8$$

Result for minimization of total transmission loss with ZIP load1 is shown in Table 2. The best case is found to be case 5 (with 4 WT-based DGs) with the maximum reduction in the transmission loss. The voltage, PLT and WT-based DG size, and location are shown in Figs. 5, 6 and 7, respectively. As observed from the Fig. 5, the voltage profile improves with DG and there is considerable reduction in the transmission loss. With more number of DGs, there is no significant improvements in the results.

Table 2 Result for minimization of total transmission loss with ZIP load1

	Case 1: Without DG(WT)	Case 2: With 1 DG(WT)	Case3: With 2 DG (WT)	Case 4: With 3 DG (WT)	Case 5: With 4 DG(WT)
PLT(p.u.MW)	0.2878	0.2734	0.2606	0.2524	0.2479
QLT(p.u.MVar)	-2.9963	-3.1542	-3.2776	-3.3612	-3.3831
Minimum voltage	0.9669 (At bus 4)	0.9676p.u. (At bus 4)	0.9662p.u. (At bus 4)	0.9659p.u. (At bus 4)	0.9645p.u. (At bus 4)
Total active load (p.u.MW)	28.5763	28.5679	28.55421	28.5317	28.5130
Total reactive load (p.u.MVar)	5.8155	5.8138	5.8085	5.8064	5.8026
% age reduction in loss		5 %	9.45 %	12.30 %	13.86 %
Optimal bus location of DG (WT)		3	3,10	3,5,10	3,5,6,10
Total DG(WT) size (p.u.MW)		0.3254	0.6508	0.8679	0.9527
Total DG(WT) size(p.u.MVar)		0.0464	0.0564	0.0664	0.0664

Fig. 5 Voltage profile with and without WT-based DG with ZIP load1

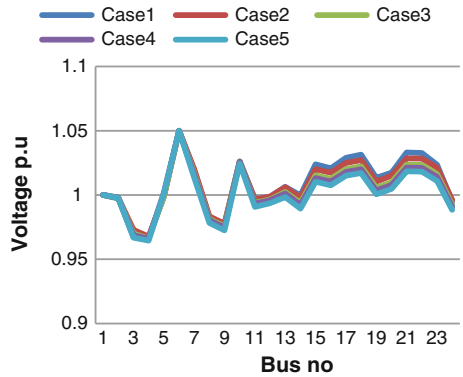


Fig. 6 Total active power loss with and without WT-based DG with ZIP load1

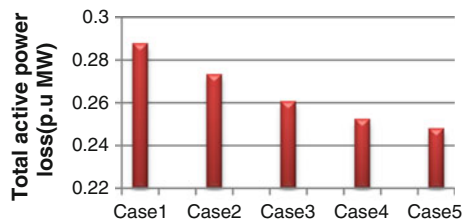
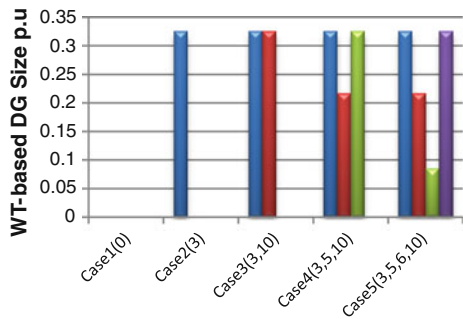


Fig. 7 WT-based DG size in p.u with ZIP load1



5.3 Results with ZIP Load2

Table 3 shows the results for all the cases with ZIP load2. The ZIP load2 has ZIP load coefficient taken as:

$$A_p = 0.2, B_p = 0.2, C_p = 0.6, A_q = 0.2, B_q = 0.2, C_q = 0.6$$

The voltage, PLT and WT-based DG size, and location are shown in Figs. 8, 9 and 10, respectively. In the case of ZIP load2 also, there is improvement in the voltage profile at all the buses with integration of DGs, and there is reduction in the

Table 3 Result for minimization of total transmission loss with ZIP load2

	Case 1: Without DG(WT)	Case 2: With 1 DG(WT)	Case 3: With 2 DG(WT)	Case 4: With 3 DG (WT)	Case 5: With 4 DG(WT)
PLT(p.u.MW)	0.2880	0.2738	0.2604	0.2518	0.2469
QLT(p.u.MVar)	-2.9868	-3.1406	-3.2707	-3.3455	-3.3828
Minimum voltage	0.9655 (At bus 3)	0.9686p.u. (At bus 4)	0.9675p.u. (At bus 4)	0.9662p.u. (At bus 3)	0.9652p.u. (At bus 3)
Total active load (p.u.MW)	28.6242	28.6047	28.5570	28.5164	28.4983
Total reactive load (p.u.MVar)	5.8252	5.8213	5.8116	5.8033	5.7996
percentage reduction in loss		4.93 %	9.53 %	12.56 %	14.27 %
Optimal bus location of DG (WT)		3	3,10	3,5,10	3,5,6,10
Total DG(WT) size (p.u.MW)		0.3254	0.6508	0.8679	0.9527
Total DG(WT) size (p.u.MVar)		0.0464	0.0564	0.0664	0.0664

Fig. 8 Voltage profile with and without WT-based DG with ZIP load2

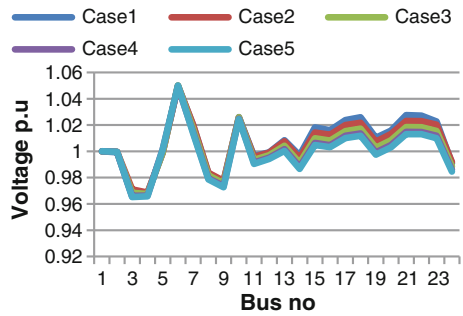


Fig. 9 Total active power loss with and without WT-based DG with ZIP load2

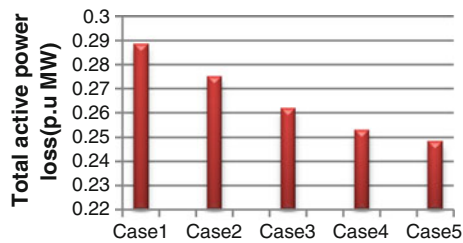
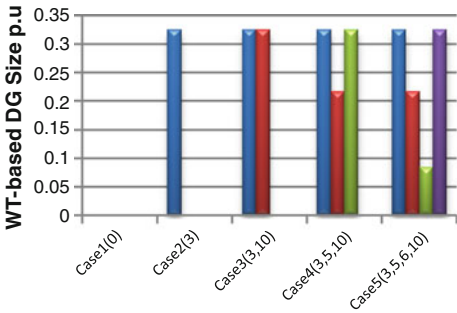


Fig. 10 WT-based DG size in p.u with ZIP load2



transmission loss as observed from the figures, however, with increase in number of DGs, there is no further improvements in the results.

5.4 Results with Variable ZIP Load

The results with the total transmission loss minimization have been determined by solving nonlinear optimization problem with variable ZIP load are given in Table 4.

Table 4 Result for minimization of total transmission loss with variable ZIP load

	Case 1: Without DG(WT)	Case 2: With 1 DG(WT)	Case 3: With 2 DG(WT)	Case 4: With 3 DG (WT)	Case 5: With 4 DG(WT)
PLT(p.u.MW)	0.2897	0.2751	0.2518	0.2532	0.2483
QLT(p.u.MVar)	−2.9647	−3.1243	−3.2519	−3.3385	−3.3633
Minimum voltage	0.9648 (At bus 3)	0.9683p.u. (At bus 4)	0.9670p.u. (At bus 4)	0.9665p.u. (At bus 3)	0.9645p.u. (At bus 3)
Total active load (p.u.MW)	28.6555	28.6301	28.5773	28.5583	28.5189
Total reactive load (p.u.MVar)	5.8317	5.8265	5.8158	5.8119	5.8039
% age reduction in loss		5.03 %	9.63 %	12.59 %	14.29 %
Optimal bus location of DG (WT)		3	3,10	3,5,10	3,5,6,10
Total DG(WT) size (p.u.MW)		0.3254	0.6508	0.8679	0.9527
Total DG(WT) size (p.u.MVar)		0.0464	0.0564	0.0664	0.0664

Fig. 11 Voltage profile with and without WT-based DG with variable ZIP load

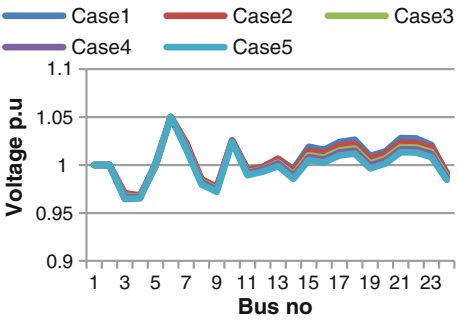


Fig. 12 Total active power loss with and without WT-based DG with constant load

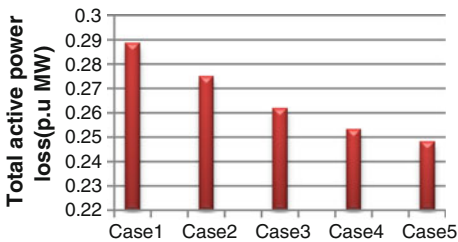
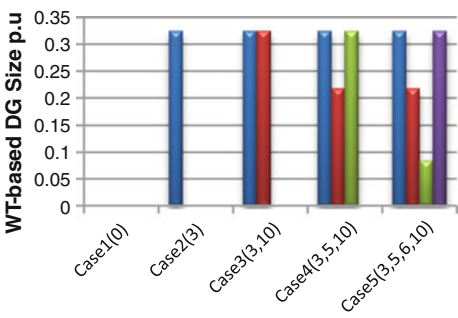


Fig. 13 WT-based DG size in p.u with variable ZIP load



Voltage profile, PLT and WT-based DG size, and location are shown in Figs. 11, 12, and 13, respectively. With variable ZIP load at all the buses, there is improvement in voltage profile at all the buses and transmissiion loss reduces considerably.

5.5 Comparison of Total Real Power Loss and Percentage Reduction in All Cases

Total real power loss in all cases is shown in Fig. 14, it is observed that in case 1 the total loss is 0.2897p.u MW with variable ZIP load and minimum loss is 0.2731p.

Fig. 14 Total real power loss in all cases with and without ZIP load

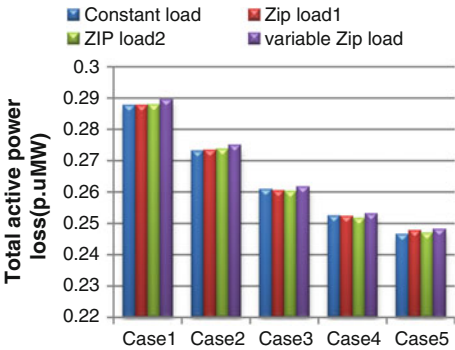
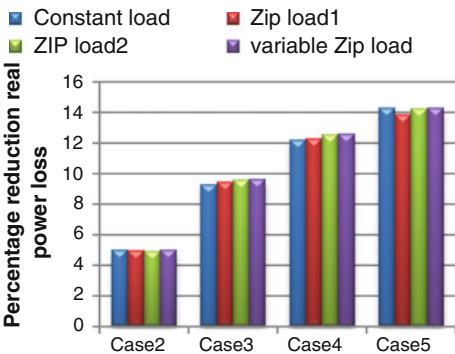


Fig. 15 Percentage reduction in losses in all cases with and without ZIP load



u MW with constant load. In case 2, the loss is 0.2732p.u MW with constant load and 0.2751p.u MW with variable ZIP load. In case 3, the loss is 0.2604p.u MW with ZIP load2 and 0.2618p.u MW with variable ZIP load. In case 4, the loss is 0.2518p.u MW with ZIP load2 and 0.2532p.u MW with variable ZIP load. In case 5, the loss is 0.2465p.u MW with constant load and 0.2483p.u MW with variable ZIP load. The total percentage reduction losses in all cases is shown in Fig. 15. It is observed that in case 2 the percentage reduction in loss is minimum 4.93 % with variable ZIP load2 and maximum is 5.03 % with variable ZIP and constant load. In case 3, the percentage reduction in loss is minimum 9.31 % with constant load and maximum is 9.63 % with ZIP load2. In case 4, the percentage reduction in loss is minimum 12.20 % with constant load and maximum is 12.59 % with variable ZIP load. In case 5, the percentage reduction in loss is minimum 13.86 % with ZIP load1 and is 14.32 % with constant load. Thus, based on the results, it is observed that with different cases of ZIP load, the losses reduction is different and with variable ZIP load at all the buses, the losses are observed slightly higher compared to other cases of ZIP load due to more reactive power requiremet for variable ZIP load. The percentage reduction in the losses increases with presence of DGs.

6 Conclusions

In this work, the wind speed samples are produced using MCS. The power outputs of different types of wind turbines are obtained. Wind turbines are integrated in the power network considering different ZIP load to observe the impact on the transmission loss. It is observed that in case of ZIP load the losses are more in the system without WT-based DG as compared to the constant load model. But when WT-based DGs are added in the system, there is reduction in active power loss in case of ZIP load as compared to the constant P, Q load. With increase in the number of WT-based DGs, the power loss reduces considerably. With more number of DG, as in case 3 and case 4, it observed that the percentage reduction in loss is more with variable ZIP load compared to constant load model. But in case 5, the percentage reduction in loss is more with constant load compared with ZIP load1. The study carried out will help the system operator in competitive electricity market to better plan the system with renewable energy sources and taking into account the better operational aspects with reduction in transmission loss and increase in the overall efficiency of the system.

References

1. Borges CLT, Falcao DM. Optimal distributed generation allocation for reliability, losses and voltage improvement. *Int J Electr Power Energy Syst* 2006; 28:413–20.
2. Lakshmi Devi A, Subramanyan B. Sizing of DG unit using genetic algorithms to improve the performance of radial distribution systems. *Int J Electron Electr Eng* 2010; 13:48–55.
3. Saedighzadeh M, Rezazadeh A. Using genetic algorithm for distributed generation allocation to reduce losses and improve voltage profile. *World Acad Sci Eng Technol* 2008; 3(7):251–6.
4. Celli G, Ghiani E, Mocci S, Pilo F. A multi-objective evolutionary algorithm for the sizing and siting of distributed generation. *IEEE Trans Power Syst* 2005; 20:750–7.
5. Moradi MH, Abedini M. A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems. *Int J Electr Power Energy Syst* 2012; 34:66–74.
6. Rotaru F, Chicco G, Grigoras G, Cartina G. Two-stage distributed generation optimal sizing with clustering-based node selection. *Int J Electr Power Energy Syst* 2012; 40:120–9.
7. Chang RW, Mithulanathan N, Saha TK. Novel mixed-integer method to optimize distributed generation mix in primary distribution systems. In: 21st AUPEC. Brisbane, Australia; 25–28 September, 2011.
8. Viral R, Khatod DK. Optimal planning of distributed generation systems in distribution system: a review. *Renew Sustain Energy Rev* 2012; 16:5146–65.
9. M.M.A. Salama, A.Y. Chikhani, A simplified network approach to the VAR control problem for radial distribution system. *IEEE Trans. Power Delivery*. 8(3) (1993).
10. Atwa YM, El-Saadany EF, Salama MMA, Seethapathy R. Optimal renewable resources mix for distribution system energy loss minimization. *IEEE Trans Power Syst* 2010; 25(1):360–70.
11. Ochoa LF, Harrison GP. Minimizing energy losses: optimal accommodation and smart operation of renewable distributed generation. *IEEE Trans Power Syst* 2011; 26(1):198–205.
12. Atwa YM, El-Saadany EF. Probabilistic approach for optimal allocation of wind based distributed generation in distribution systems. *IET Renew Power Gener* 2011; 5(1):79–88.

13. Hung DQ, Mithulananthan N, Bansal RC. Analytical expressions for DG allocation in primary distribution networks. *IEEE Trans Energy Convers* 2010; 25(3):814–20.
14. Hung DQ, Mithulananthan N. Multiple distributed generators placement in primary distribution networks for loss reduction. *IEEE Trans Ind Electron* 2013; 60(4):1700–8.
15. Kusiak A, Li W. Short-term prediction of wind power with a clustering approach. *Renew Energy* 2010;35(10):2362–69.
16. Kusiak A, Li W. Virtual models for prediction of wind turbine parameter. *IEEE Trans Energy Conversion* 2010; 25(1):245–252.
17. Abdel-Aal RE, Elhadidy M.A, Shaahid SM. Modling and forecasting the mean hourly wind speed time series using GHDM-based abductive network. *Renewable Energy* 2009; 34:1686–1699.
18. IEEE Reliability Test System, A report prepared by the Reliability Test System Task Force of the Applications of Probability Methods Subcommittee, *IEEE Trans. on Power Apparatus and Systems* PAS-98 Dec.1979; 2047–2054.

Use Case-Based Software Change Analysis and Reducing Regression Test Effort

Avinash Gupta and Dharmender Singh Kushwaha

Abstract It is very difficult for software organizations to fulfill users' requirement, as they change frequently. It is an organization's major responsibility to get rid of this change as soon as possible in order to compete in market. A brief analysis of these changes is very important before implementing in order to prove them profitable to the users. This work proposes a UML model-based approach using the use case and class diagrams for impact analysis and decision table that is applicable in early decision making and change planning. Later, by using the impact set we estimate the regression test effort required for the effected change in the software. The reduction in test effort observed ranges from 20 to 65 % saving significant software testing cost too. The proposed methodology obtains a reduction of 37.5 % on an average.

Keywords Use case • Decision table • Software change impact analysis • Test effort

1 Introduction

As enterprises grow, its need and functionalities may change. The change demanded by the user/client may be required to be done either during the design phase, or after the coding has been done or else after the software application has become functional at the client side. The proposed approach is based on the use case and class diagrams of the software under consideration [1]. The use case diagram with descriptions is used in SCIA. A software system may contain one or more use cases. Each use case has its main flow of event and alternate flow of event. Use cases can be later transformed into class diagrams. The class diagram is also

Avinash Gupta (✉) · D.S. Kushwaha
Department of Computer Science and Engineering, MNNIT Allahabad,
Allahabad, India
e-mail: avinashg.mnnit@gmail.com

D.S. Kushwaha
e-mail: dsk@mnnit.ac.in

used in this proposed approach to identify the methods (and hence the concerned classes) that need to be modified in order to arrive at impacted classes. From the impacted classes, the concerned test cases are selected for regression testing. This reduces the amount of test cases that has to run after changes/modifications have been incorporated in the software.

2 Related Work

There are various strategies for performing SCIA. These strategies are based on some parameters which are considered during the requirements engineering process or the development phase [2]. SCIA approaches can be broadly classified as automatable and manual [3]. Manual approaches are best performed by human beings. These approaches require fewer infrastructures but may be harder in their impact estimation than the automatable ones. SCIA approach often employs algorithmic methods in order to identify change impact [4]. Requirements dependency webs and object models are examples of structured specification. Bengtsson and Bosch [5] employ statistical meta-analysis techniques to investigate the ability of object oriented (OO) metrics to predict change-proneness of a system. Some automated tools of impact analysis are Rational rose [6] and Visual paradigm [7].

Minhas and Zulfiqar [8] investigate the role of understanding code changes during software development process, and also explores the developer's information needs for understanding changes and their requirements for the corresponding tool support. SCIA approach in this work is based on information retrieval process as discussed in [9] that is used to derive the information contained in use case flow events, with respect to requested change. Other approaches [10] discuss about deriving effort from requirement and analysis phase. Kushwaha and Misra [11] proposes a technique for estimating the test effort and establishes cognitive information complexity measure (CICM) as an appropriate estimation tool. Authors in [12, 13] propose novel methods for software change management and its impact analysis. Khurana et al. [14] propose a technique for change impact analysis and its regression test effort estimation. The aim is to reduce the number of test cases that have to be rerun.

3 Proposed Work

3.1 *Software Change Impact Analysis*

The change demanded by the user/client may be required to be done either during the design phase or after the coding has been done or else after the software application has become functional at the client side. The proposed approach is based on the use case and class diagrams of the software under consideration. The use case diagram with descriptions is used in SCIA. A software system may contain

one or more use cases. Each use case has its main flow of event and alternate flow of event. Use cases can be later transformed into class diagrams. The class diagram is also used in this proposed approach to identify the methods (and hence the concerned classes) that need to be modified in order to arrive at impacted classes. From the impacted classes, the concerned test cases are selected for regression testing. This reduces the amount of test cases that has to run after changes/modifications have been incorporated in the software.

Dependency of use cases are utilized in the proposed approach. The proposed approach includes the following steps:

- *Read SCRF* Users or Developers, who request changes in the existing software system, fills the software change request form (SCRF) as per new/changed requirement. After getting the SCRF filled, controller configuration tool (CCT) analyzes the SCRF. The important field of the SCRF, i.e., the change requested field is stored in a new directory.
- *Parsing and Extraction* In this phase, change request file is parsed. Parser parses the stop words like is, are, am, this, that, etc., from the stop word file. The parsed keywords are stored in an output file.
- *Impacted Use Cases* This phase is the information retrieval (IR) phase. After parsing the SCRF for all extracted keywords, the flow of events directory provides the impacted use case with respect to each keyword. This process recursively runs for all keywords. For each flow of event file, respective use case name is stored.
- *Check Similar Use Cases* In this step similar use cases names are checked.
- *Delete Similar Use Cases* In the previous step, it might be possible that there may be some redundant use cases. In this phase, these redundant use cases are removed to avoid the ambiguity.
- *Final Impacted Use Cases* The final outcome of this step is the name of impacted use cases.

Proposed scheme is illustrated by the ten open sources Java-based and some self-made projects, references of each project is shown in Table 1.

3.2 Use Case Conversion to Decision Table

An algorithm has been proposed for fetching a use case in a particular template from a file and then converting this use case into a decision table. This decision table is further optimized so as to obtain reduced set of regression test cases that need to be executed once the requested changes have been incorporated in the application software. The code takes the input as use case and seeks for the condition action clauses. Recording of these clauses is done and distinct conditions and actions are put into separate hash map and then correspondingly a combination of conditions performs a particular set of actions. This is recorded in a table (Decision table).

Table 1 Project references

Project name	References
ATM	http://www.math-cs.gordon.edu/courses/cs211/ATMExample
Coffeemaker	http://agile.cse.ncsu.edu/SEMaterials/tutorials/coffee-maker
Address book	http://www.javabeginner.com/java-swing/java-swing-address-book
Library management	http://www.projectsparadise.com/library-information-system-java
Airline booking	http://www.muengineers.in/computer-project-list/java-projects-list
Pharmacy	http://www.projectsparadise.com/pharmacy-management-system-java
Payroll	http://www.projectsparadise.com/payroll-management-java
Shipment	Self-made
Java operation	Self-made
Student record information	http://www.projectsparadise.com/student-record-information-system-java

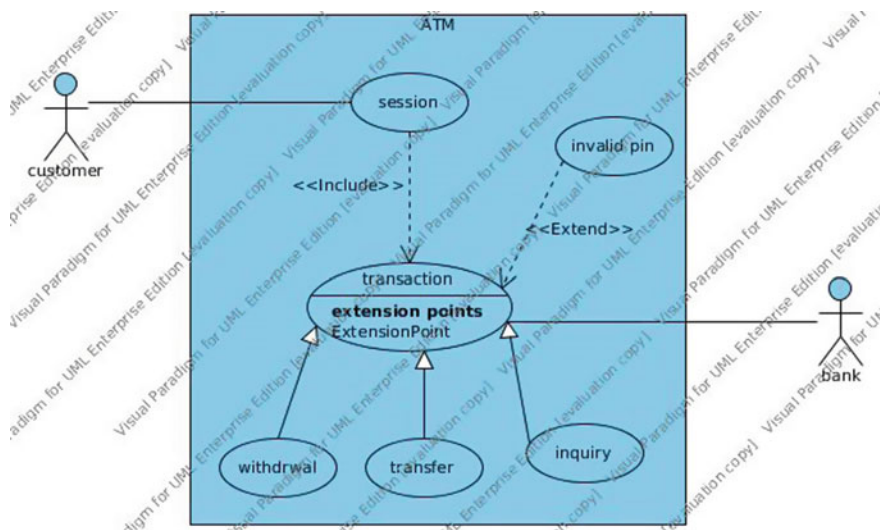


Fig. 1 Use case diagram designed by visual paradigm tool

The decision table thus generated is given as an input to the optimizer code which optimizes the decision table and percentage reduction is calculated (Fig. 1).

Here is the input (Use case)

Name: ATM System

Actor: User/Customer

Main Success Scenario

- Step 1: Insert Card.
- Step 2: Validate Card and asks for PIN.

Table 2 Decision table derived from use case

Conditions																
Card_Invalid	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y
PIN_Invalid	N	N	N	N	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y
PIN_Invalid_thrice	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
User_not_authenticated	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Actions																
Reject_the_card	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y
No_cash_withdrawl	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ask_for_retry	N	N	N	N	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y
Block_the_card	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y

Step 3: Enter PIN.

Step 4: Validate PIN.

Step 5: Allows access to account.

Else

- (1) Card Invalid—Reject the Card and No Transaction allowed.
- (2) PIN Invalid thrice—Block Card and No Transaction allowed.
- (3) User not authenticated—No Transaction allowed.

Table 2 illustrates the decision table for this use case before and after the optimization. Table 3 illustrates the optimized decision table. This verifies the reduction in test cases required after certain modification is incorporated in existing software.

Once the use case is transformed into a decision table, each of such row demands a test case to be written. Later, through the previously proposed decision table optimization algorithm, the entries of the tables are analyzed and reduced to remove the redundant conditions that are present resulting in minimum number of test cases required to test the application under consideration.

Table 3 Optimized decision table

Conditions											
Card_Invalid	N	N	N	N	N	N	N	Y	Y	Y	Y
PIN_Invalid	N	N	N	Y	Y	Y	N	N	Y	Y	
PIN_Invalid_thrice	N	N	Y	N	N	Y	N	Y	N	Y	
User_not_authenticated	N	Y	–	N	Y	–	–	–	–	–	–
Actions											
Reject_the_card	N	N	N	N	N	N	Y	Y	Y	Y	
no_cash_withdrawl	N	Y	Y	N	Y	Y	Y	Y	Y	Y	
Ask_for_retry	N	N	N	Y	Y	Y	N	N	Y	Y	
Block_the_card	N	N	Y	N	N	Y	N	Y	N	Y	

4 Performance Analysis

We demonstrate the proposed methodology using a small self-made mini-application computer operations tutorial (COT) on Java that takes a mathematical expression as input, parses it, and displays the solution as the output (Fig. 2).

For a change requested in equality classes, in the existing test suite there are 190 test cases, out of which we require to rerun only 86 test cases, thus required regression testing effort [15] for the change will be 0.46E where E is the effort required to run the existing test suite and the % of average reduction in testing effort will be 54 % of the existing system testing effort.

Figure 3 shows the relation between impact area of a class and reduction in effort during regression testing. As the number of impacted classes rises for a class, more the effort is required for regression testing.

Fig. 2 Class diagram for COT

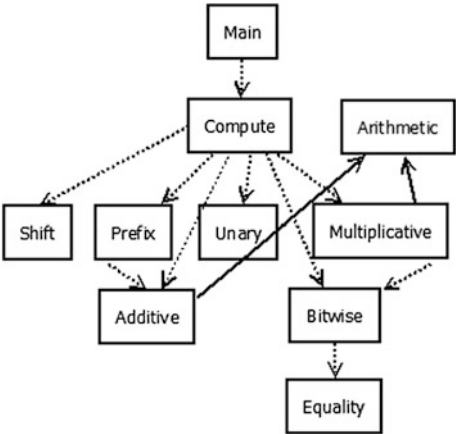
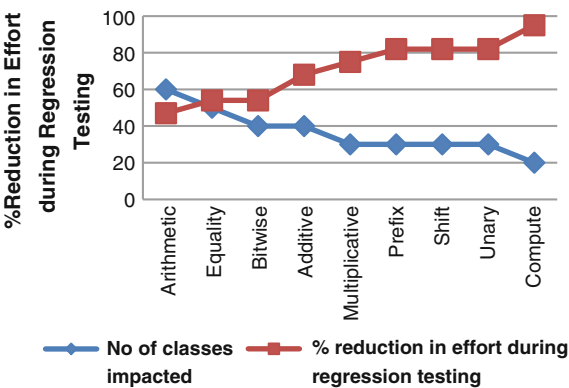


Fig. 3 Relation between impact area of a class and reduction in effort during regression testing



5 Conclusion

Software change impact analysis (SCIA) needs to be carried out for every software change request. This shall result in understanding the risk associated with several critical software engineering tasks such as time, cost, and effort estimation along with the regression testing. The proposed approach is more efficient in terms of level of abstraction, since the proposed approach works with use cases that are later mapped to classes in class diagram. The reduction in test effort observed ranges from 47 to 95 % saving significant software testing cost.

Once the use case is transformed into a decision table, each of such row demands a test case to be written. Later, through the previously proposed decision table optimization algorithm, the entries of the tables are analyzed and reduced to remove the redundant conditions that are present resulting in minimum number of test cases required to test the application under consideration. Here, we observe a reduction of 37.5 % on an average.

References

1. Aprna Tripathi, D.S. Kushwaha and Prof. Arun Misra, "Quality Validation of Software Design before Change Implementation through Complexity Measurement", In AISC series of Springer-Dec 2013.
2. H. Lu, Y. Zhou, B. Xu, H. Leung, and L. Chen. The ability of object-oriented metrics to predict change-proneness: a meta-analysis. *Empirical Softw. Engg.*, 17(3):200–242, June 2012.
3. H. Kagdi and D. Poshyvanyk. Who can help me with this change request? In *Program Comprehension, 2009. ICPC'09. IEEE 17th International Conference on*, pages 273–277. IEEE, 2009.
4. Aprna Tripathi, Dharmender Singh Kushwaha, Arun Kumar Misra, "Analysis of Impacted Classes and Regression Test Suite Generation", 2nd International Conference on Advance Computing and Creating Entrepreneurs (ACCE 2013), Udaipur, February 2013.
5. P. Bengtsson and J. Bosch. Architecture level prediction of software maintenance. In *Software Maintenance and Reengineering, 1999. Proceedings of the Third European Conference on*, pages 139–147. IEEE, 1999.
6. Rational rose. <http://www-03.ibm.com/software/products/us/en/ratirosefami/>. [Online; accessed 03-June-2015].
7. Visual paradigm. <http://www.visual-paradigm.com/>. [Online; accessed 03-June-2013].
8. N. Minhas and A. Zulfiqar, 'An Improved Framework for Requirement Change Management in Global Software Development', *Journal of Software Engineering and Applications*, vol. 07, no. 09, pp. 779–790, 2014.
9. R. S. Pressman. *Software Engineering: A Practitioner's Approach*. McGraw-Hill Higher Education, 5th edition, 2001.
10. Ashish Sharma, Manu Vardhan and Dharmender Singh Kushwaha, "A Versatile Approach for the Estimation of Software Development Effort Based on SRS Document", *International Journal of Software Engineering and Knowledge Engineering (IJEKE)*, Volume 24, Issue 01, February 2014, pp. 1–42.
11. D.S. Kushwaha and A.K Misra, "Software Test Effort Estimation", *ACM SIGSOFT Software Engineering Notes*, Vol. 33, No. 3, May 2008.

12. Aprna Tripathi, Dharmender Singh Kushwaha, Arun Kumar Misra, "Software Change Validation Before Change Implementation", 2nd International Conference on Advance Computing and Creating Entrepreneurs (ACCE 2013), Udaipur, February 2013.
13. Rajat Swapnil, Aprna Tripathi And Dharmender Singh Kushwaha, "Software Change Validation Using Class Diagram and SRS", 3rd IEEE International Advance Computing Conference (IACC-2013), February 2013.
14. Prateek Khurana, Aprna Tripathi And Dharmender Singh Kushwaha, "Change Impact Analysis and its Regression Test Effort Estimation", 3rd IEEE International Advance Computing Conference (IACC-2013), February 2013.
15. Prateek Khurana, Aprna Tripathi And Dharmender Singh Kushwaha, "Change Impact Analysis and its Regression Test Effort Estimation ", 3rd IEEE International Advance Computing Conference (IACC-2013), Ghaziabad, February 2013.

Tackling Supply Chain Management Through RFID: Opportunities and Challenges

Prashant R. Nair and S.P. Anbuudayasankar

Abstract RFID is a powerful technology that provides enterprises with improved inventory tracking, transparency, and visibility, thereby enhancing operational efficiency and better engagement and dialog with shipping channels and storage depots. RFID helps enterprises to attain just in time (JIT), vendor managed inventory (VMI), or zero inventory levels. This paper explores the current state of affairs with respect to technology of RFID and its current usage in diverse application domains and contexts of supply chain. Information is vital for decision support and the quality of this information is emerging as the most vital metric to measure supply chain performance. RFID provides supply chain managers with accurate, actionable, and timely information for decision support. Various implementation barriers and RFID adoption challenges are discussed. Successful demonstrations of RFID usage for supply chain activities by progressive companies are showcased. By providing a glimpse into the immense business value and wide-ranging strategic advantages of using RFID for supply chain management, the paper exhorts companies to tap the vast potential of this exciting and promising technology. Integration of RFID within emerging trends like Internet of things and social media analytics and cloud (SMAC) technologies is the future research direction of this technology.

Keywords RFID · Supply chain management (SCM) · Inventory · Tag

P.R. Nair (✉)

Department of Computer Science & Engineering, Amrita School of Engineering Coimbatore,
Amrita Vishwa Vidyapeetham, Amrita University, Amrita Nagar P.O,
Coimbatore 641112, India
e-mail: prashant@amrita.edu

S.P. Anbuudayasankar

Department of Mechanical Engineering, Amrita School of Engineering Coimbatore,
Amrita Vishwa Vidyapeetham, Amrita University, Amrita Nagar P.O,
Coimbatore 641112, India

1 Introduction

The precursor technology to RFID, the bar code, no doubt, made a lasting impression in the retailing industry. Although, it improved operational efficiency as an Auto-ID technology, there were some disadvantages. The bar code is not in a position to distinctively recognize the specific object and its attributes, the production date of the specific items, the lot of the items was made, and the expiration date of the items. These issues were more or less rectified with the advent of RFID. The basic premises of both these auto-ID technologies are similar, i.e., to give item identification. RFID and bar code differ in their method of reading data. In RFID, a tag is read by the RFID reader using radio frequency signals, while a printed label is read by the bar code, which uses optical laser or imaging technology.

Inventory tracking, visibility, and transparency up to the item level are important for supply chain planning and execution. This visibility improves operational efficiency, reduces operating costs and channel volume. It also enhances supply chain forecasting and planning capabilities [1]. The purpose of RFID is to allow data to be transmitted by a portable device, called a tag, which is read by a reader and processed as per the needs of various applications. The data transmitted by the tag may provide identification or location information, or specifics about the product tagged, such as color, price and date of procurement [2] (Fig. 1).

RFID tag is basically nothing but a microchip connected to a tiny antenna. The microchip is empowered to capture a definite quantity of data. Frequency of radio waves through which communication and dialog between the reader and tag happen can be in a wide range from microwave (6 GHz) to very low (125 kHz). The tangible benefits of using RFID for SCM include supply chain inversion, better regulation, high cost benefit, communication workflow, and visibility [3]. Supply chain inversion refers to the transformation of a push system to pull. Use of RFID improves inventory visibility and reduces lead time in the supply chain. RFID helps enterprises to comply with various dynamic and every changing government legislations and regulation on safety and public information on products. RFID adds to the bottom-line by shrinking labor costs and this provides a tangible benefit if you

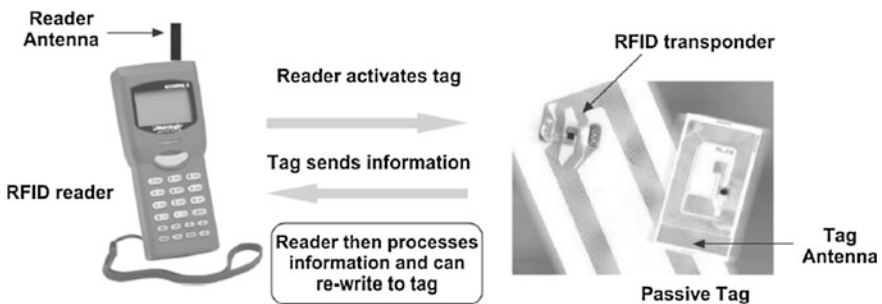


Fig. 1 RFID communication workflow [2]

consider a cost benefit analysis. Various real-time RFID applications include identity authentication [4] and shrink reduction and avoidance. The usage of electronic sealing utilizing RF tags can validate the authenticity and point of origin of a product. RF tags can be used to prevent spurious shrinks as also malicious or non-malicious shrink. Theft and fraud usually result in malicious shrink, whereas non-malicious shrink is connected to inefficient product handling. Tags have the advantage of reusability as well as they can be deployed in new products as well. Another option is to retain it in the product so as to have a secure serial number at all times [5]. All stakeholders like factories, suppliers, shippers, wholesalers, or retailers are advantaged from inventory tracking and visibility. The retail space has been the primal driver of RFID usage for SCM. Scottish Courage Brewing Ltd, which has captured almost 50 % of draught beer market in United Kingdom, has made substantial RFID investments of US \$14 million to track their 2 million kegs and 736 transport vehicles [5]. RFID has impacted this company by reducing keg losses and rendering the delivery process more efficient and effective. The company claims that their cycle times have been improved by 4 days after the deployment of RFID. HR Group, one of EU's largest shoe companies with presence in 20 countries and comprising of 750 stores, has embedded RFID chips in each of their shoes across all stores. RFID is used to track shipments from point of sale to warehouses and factories some of whom are situated in Asia. This has also helped the company reduce theft and counterfeiting [5]. The mega retailer, Walmart who stocks more than 500,000 items has mandated suppliers to use RFID tags as early as 2003 [4].

2 RFID Application Areas

RFID applications have been so widespread in various sectors and verticals that it has been billed as one of the revolutionary technologies of the future technology landscape. It is widely used in several contexts: prominent among them retailing, secured entry to office spaces, manuscript tracking, cattle or domesticated animal identification, automobile access, pay-at-the-pump gasoline sales, artifact validation, sports timing, and wireless payment [6].

The total RFID market which includes its building blocks like tags and readers has been seeing exponential growth and traction for the last 10 years. This market is estimated to be north of US \$ 25 billion by end of this year. This spectacular spurt will be enabled by brisk fall in the prices of RFID tags and widespread use across various verticals. One estimate suggests that the RFID tag market will grow to the order of trillions of tags by end of 2015 [7]. Almost 1.5 million employment opportunities primarily in the retail space [8] is yet another expected outcome of the RFID boom by 2016. RFID market will intersect with complementary and associated technologies like smart phones, GPS, Wi-Fi, and WiMax, all of which may be grouped as radio frequency (RF) technologies [9]. The promise of RFID is entrenched in the power of information, which propels the knowledge economy and transforms our society today.

RFID Applications are wide-ranging as evidenced by the following:

- US Department of defense, Walmart and Target are some of the large players who have mandated all their major suppliers to switch to RFID [10].
- The WHO wants to use RFID for checking the flow of counterfeit drugs.
- The mobil speed pass system which has become very popular in developed economies, where one can pay for gas by placing a tag in an RFID system.
- Patients with Alzheimer's disease and similar memory illnesses could make use of items with RFID tags to perform various activities [11].
- Theme parks like Disney, where there are thousands of footfalls on a daily basis are providing wrist bands with RFID to children to track them even if they are separated from their guardians due to overcrowding of the park [12].
- Casinos are embedding RFID in gambling chips for extending their visibility for tracking table play.
- The Chinese government has affixed RFID tags on its exotic national animal, pandas, as endangered and exotic species.

The advantages that enterprises can achieve by adopting and deploying RFID technologies include [6]:

- Improved visibility into requirements of the customer.
- Effective transparency, tracking, and visibility of inventory.
- Efficient business processes.
- Dependable and precise demand forecast.
- Better ROI.
- Improved productivity.
- Lesser operating costs.
- Superior productivity.
- Improved counterfeit identification and theft prevention.
- Better response and engagement with customer.

3 RFID for SCM

The ICT innovation called RFID is showing tremendous promise as key SCM enabler. RFID is a powerful technology that provides enterprises with improved inventory tracking, transparency, and visibility, thereby enhancing operational efficiency and better engagement and dialog with shipping channels and storage depots. RFID helps enterprises to attain just in time (JIT), vendor managed inventory (VMI) or zero inventory levels. RFID also facilitates better decision support for operational managers at the assembly or shop floor by providing real-time, precise and actionable information. It has the potential to cut costs, reduce shrinkage and spurious items, and increase sales by reducing out-of-stocks.

All stakeholders in a supply chain like manufacturers, retailers, and distributors are impacted by the use of RFID as a tagging technology for inventory tracking.

A research project of the University of Arkansas showed stores having tangible benefits and savings of 16 % using RFID. This is primarily by efficient replenishment of items and avoids upsetting customers who become unhappy when they do not get the products of their choice. This is no pushover when one considers the fact that Walmart estimates a loss of US \$ 1 billion due to this trend [13].

RFID technologies as well as technologies like decision support systems and software agents are universal infrastructure for mobile commerce and introduce process freedom [14] for various cogs in the supply chain wheel. Retail giants like Walmart and Target and also the US Department of Defense are early RFID adopters for SCM operations especially inventory tracking and management [6]. Walmart has made it mandatory for its entire top suppliers like HP in the electronic good vertical to embed RFID in pallets and units. HP, thereby tweaked and customized their transportation processes to accommodate RFID.

Most of the top retailers in the Fortune 500 list are in various stages of adoption and migration to RFID-based tagging. Ford uses RFID for component replenishment and tracking of shippers [10]. Examples of some progressive enterprises that have actively deployed RFID for SCM are DHL, YCH in Singapore, Dolomiti Superski in Italy, McCarran Airport in Las Vegas, and NHK in Japan [15].

Tags placed in oil and gas pipelines will make maintenance and monitoring easier. Hospitals will be to track the whereabouts of lifesaving equipment. Pharmaceutical companies are able to reduce counterfeit or spurious drugs and pills. Aerospace industry can effectively dispose and handle hazardous supplies. Sea port security will be improved; the logistics and transportation industry will have efficient control and regulation of all transport hubs.

Even though RFID has been in vogue for quite some time, usage of RFID for SCM is relatively recent and not very widespread. There is a misconception that RFID is only a replacement technology over bar codes. There seems to be lack of awareness on the multifarious RFID benefits and thereby companies do not tweak their business and SCM processes based on RFID capabilities. Benefits include reduction of human labor from several workflows, improvement of the bottom-line, and business value creation for enterprises.

4 RFID Applications in SCM Processes

This paper explores of the usage of RFID in the following SCM processes.

4.1 Efficient Management of Inventory and Asset Tracking

Efficient tracking and management of inventory and asset tracking is by far the most prominent and prevalent RFID intervention in SCM. RFID is used to authenticate and track the identification of goods at the unit, pallet, case, and carton levels. It would be

unnecessary to open cartons. RFID facilitates unit-level tagging of various items and will be the mainstays till the retailers switch completely to automated check outs. However, this trend may not catch up everywhere in the near future.

RFID tags have unique serial identifier information that links each lot with a matching bill of lading sent from the point of origin. Because RFID readers can scan tags many times during a one second period, the serial identifier the application making the data request from getting repeated counts of the same objects [15]. The accurate inventory information is the key for improving the performance of the supply chain. RFID technology enables stakeholders to precisely track and locate every pallet through any cycle in the supply chain and making dynamic and flexible routing decisions as per need and circumstance [16].

One study estimates that the losses in the US retail sector due to poor inventory visibility is a whopping \$ 70 billion annually [16]. RFID-enabled inventory management would reduce this to a great extent. It would be possible to reduce counterfeiting and also track goods that are difficult to track like beer kegs, the application developed by TrenStar. Proctor & Gamble, after RFID deployment for SCM estimates savings of US \$200 million in inventory-carrying costs [17]. Logistics and supply chain service providers would be in a position to continuously monitor and track their cargo throughout the journey from source to destination. RFID can be deployed in containers, yards, factories, or even warehouses, which is done by Amcor [16]. RFID also provides a security blanket.

4.2 Vendor Managed Inventory (VMI)

VMI is something every retailer desires. Retailers can reduce manpower, storage costs, and out-of-stocks as a result of VMI. The supplier has a handle on inventory and total control due to his access to point of sale and stock information of the retailer. The responsibility of inventory management vests with the supplier rather than the retailer. There would be reliable data for demand planning and shaping using RFID. This would include stock, work-in-progress, and finished goods.

4.3 CRM

RFID usage helps to build a better relationship and engagement with the prized customer. Customers need not be upset at not seeing their favorite items and products being out-of-stock. Most consumer behavior studies have pointed out that this effect on the customer can be disastrous. With control of inventory migrating from retailer to supplier, CRM is much better. The chances of items being sent to wrong locations are also minimized herewith. This would result in reduced cost as well as labor.

Suppliers will be able to handle recall of defective goods and products more effectively using RFID. Every product or good is tied to a particular sale or return and this is logged by electronic security marker (ESM) in RFID [17]. Retailers can also offer better after-sales service. Manufacturers are also shielded in case of return by fraudulent means. The brand of both the supplier and retailers is enhanced due to the RFID tagging of the items in the reverse logistics.

4.4 Production and Manufacturing Workflow

RFID can be used for shop floor process automation. This impacts visibility of goods in the supply chain and velocity in workflow. The quality of products in the manufacturing assembly line can be monitored using RFID. Real-time data collection helps quality control departments in manufacturing companies. Tags can monitor things like pilferage, tamper, and environmental parameters such as temperature and bacterial levels. The US army is deploying RFID tags in conjunction with sensors to monitor environmental parameters especially in areas where they are having substantial transfer of goods and services [16].

5 RFID Adoption Issues

RFID adoption comes with its share of challenges and barriers. Before deployment of RFID, enterprises would have to do its homework in significant upfront planning and testing to embrace and deploy this for SCM processes and activities at both planning and execution.

One saving grace is the rapid reduction in costs of the RFID tags. In spite of this, many enterprises especially manufacturers are concerned about the high initial investment for RFID. In some cases, big retailers have mandated this to their suppliers. However, all enterprises should consider the bigger picture in terms of the tremendous medium-term to long-term benefits accrued due to the RFID usage in terms of better inventory management, reduction in costs and labor, and security.

The development of standards for encoding information on RFID tags will be critical to popularizing the technology for SCM. Presently, EPCglobal Inc., the standards body that manages Universal Product Code (UPC) information in bar codes, sets the standards for how basic product information is encoded in the RFID chips [17]. However, there are issues when new applications from different verticals log onto this technology. Standards will need to continuously evolve to keep pace with the advances in RFID.

RFID is not bereft of technology issues like radio interference. Once in a while, readers fail to read, when the tags are on liquids or metals. Some reports of interoperability with older RFID products have also been observed. Different SCM stakeholders like customers, suppliers, vendors, and third party logistic providers

use different systems for inventory control, storage, and shipping and this leads to interoperability and portability issues. Notwithstanding these issues, RFID holds tremendous promise in terms of integration with emerging game-changing technologies and applications like Internet of things [18], social, media, analytics and cloud (SMAC) stack [19], and distributed manufacturing. Integration of RFID within these trends would open new vistas, applications, and the future research directions. In addition to retail, RFID for SCM is poised to make a mark in the fashion industry [20] as well as all application areas of SCM [21].

6 Conclusion

RFID provides enterprises with improved inventory tracking, transparency, and visibility, thereby enhancing operational efficiency and better engagement and dialog with shipping channels and storage depots. As a result, RFID is helping enterprises to attain just in time (JIT), vendor managed inventory (VMI) or zero inventory levels. All these lead to considerable savings and business value for corporations. RFID application areas in various SCM processes like inventory management, asset tracking, VMI, demand shaping, production workflow, and customer relationship management are studied. Successful deployments in various companies are showcased. Various RFID adoption and implementation bottlenecks are explored. For success, enterprises should reengineer their business workflows based on RFID's capabilities with a more strategic perspective and recognize that ROI should be considered from a medium to long-term perspective. Integration of RFID within emerging trends like Internet of things and social media analytics and cloud (SMAC) technologies is the future research direction of this technology.

References

1. D'Avanzo, R., Starr, E., Von Lewinski, H: Supply chain and the bottom line: a critical link. Outlook: Accenture. 1, 39–45 (2004).
2. Nair, Prashant R.: RFID for Supply Chain Management. CSI Comm. 36(8), 14–18 (2012).
3. De Wachter, H. Pleysier: Strategic Project RF-Tags, RF-Tags for Smart Logistics. Flanders Institute for Logistics, Antwerpen (2004).
4. Coronado, A.E., A.C. Lyons, Z. Michaelides, D. F. Kehoe: Automotive supply chain models and technologies: A review of some latest developments Information, www.emeraldinsight.com/1741-0398.htm (2004).
5. Schneider, M.: Radio Frequency Identification (RFID) Technology and its Applications in the Commercial Construction Industry. University of Kentucky thesis information (2003).
6. Attaran, M.: RFID: an enabler of supply chain operations. Supp Chn Mgmt. 12(4), 249–257 (2007).
7. IDTechEx Information, www.idtechex.com/products/en/articles/00000169.asp.
8. Nair, Prashant R.: E-Supply Chain Management using Software Agents. CSI Comm. 37(4), 13–16 (2013).

9. Wyld, D. C.: RFID 101—the next big thing for management. *Mgmt Res News*. 29(4), 154–173 (2006).
10. Spekman, R. E., P. J. Sweeney: RFID: from concept to implementation. *Int JI of Phy, Dist and Logst Mgmt*, 36(10), 736–754 (2006).
11. Want, R.: RFID: a key to automating everything. *Scient Amer*, 290(1), 56–66 (2004).
12. Dignan, L.: Riding radio waves. *Baseline* 2004, 22–24.
13. RFID Journal Information, <http://www.rfidjournal.com/article/print/2314>.
14. Nair, Prashant R., Balasubramaniam, O.A.: IT Enabled Supply Chain Management using Decision Support Systems. *CSI Comm*. 34(2), 34–40 (2010).
15. E-week Information, <http://www.eweek.com/c/a/Mobile-and-Wireless/RFID-Reshapes-Supply-Chain-Management/>.
16. Michael, K., McCathie, L.: The Pros and Cons of RFID in Supply Chain Management. In: *International Conference on Mobile Business*, pp. 623–629 (2005).
17. Sabbaghi, A., Vaidyanathan, G.: Effectiveness and Efficiency of RFID Technology in SCM—Strategic Values and Challenges. *J. of Theo and Appl Electron Comm Res*. 3(2), 71–81 (2008).
18. Cerasis Information, <http://cerasis.com/2015/05/04/supply-chain-management-trends/>.
19. Nair, Prashant R.: The SMAC effect towards adaptive supply chain management. *CSI Comm*. 38(2), 31–33 (2014).
20. Wong, W.K., Guo, Z.X.: *Fashion Supply chain management using RFID technologies*. Woodhead Publishing (2015).
21. Florea (Ionescu), A.I.: Benefits and Drawbacks in using RFID (Radio Frequency Identification) System in Supply Chain Management. In: H.A. Le Thi et al (eds.) *AISC*, vol. 360, pp. 177–188. Springer, Heidelberg (2015).

Quality Improvement of Fingerprint Recognition System

Chandana, Surendra Yadav and Manish Mathuria

Abstract It is a great achievement for us to digitally store and match digital fingerprint to recognize human. But still there are some significant modifications required to improve the quality of fingerprint recognition system's result. Most of the digital fingerprint image contains noise. This unwanted information can be removed using preprocessing by image enhancement which includes various types of filtering techniques and binarization. Then the postprocessing operation which includes the computations of minutiae points and minutiae matching operation is performed. The research work is analyzed by comparing the output of filters. It exposed that unsharp filter gives more output values as compared to other filters that is better for blur fingerprint images.

Keywords Fingerprint recognition system • Enrollment • Fingerprint verification • Fingerprint identification

1 Introduction

The fingerprint images are used for person identification to identify or to contact a single person. To matches the fingerprint images, a special system is used known as the fingerprint recognition system. Generally password or identification card are used to identify the person in a particular area for accessing the system but this limitation of a particular area can be easily removed through the biometrics. Biometrics is a technology used for measuring the life characteristics because the

Chandana (✉) • Surendra Yadav • Manish Mathuria
Department of Computer Science & Engineering, Maharishi Arvind Institute
of Engineering & Technology, Sirsi Jaipur, India
e-mail: chandana_rahar@outlook.com

Surendra Yadav
e-mail: syadav66@yahoo.com

Manish Mathuria
e-mail: manish.4598@gmail.com

biometrics data such as finger print image are unique for every person. Biometrics system such as fingerprint recognition system (FRS) are reliably capable of analyzing two fingerprint images that is original and another one is imposter stored in a database using the fingerprint matching systems.

The FRS can be categorized into three sub domain: 'Fingerprint Enrollment,' 'Fingerprint Verification,' and 'Fingerprint Identification.' And last one that is a different approach from the general approach by researches, the fingerprint recognition here is referred as 'Automatic Fingerprint Recognition System' (AFRS) which is a program based.

Fingerprint Enrollment When first time an individual person uses a biometrics system, it is defined as enrollment in the device. While enrollment process occurred, then biometrics information from an individual is captured and stored in database for further operation.

Fingerprint Verification In the fingerprint database, there is one to one comparison of a captured fingerprint image which is stored with specific templates in order to verify the authenticity of individual person by his fingerprints. For all users the reference model and some templates are matched to create the authorized and unauthorized score and threshold value is estimated. For verification process positive recognition is used to prevent from multiple people using the same identity.

Fingerprint Identification Identification is done by comparing one fingerprint to many others. It is performed to establish the identity of an unknown called identification mode. It is clear that identification mode is used for both one for 'Positive Recognition' (where user need not to provide template information) as well as for the 'Negative Recognition.'

2 Prior Work

'M'ALI H.A. and NE'MA B.M.' use the fingerprint images for multiple code generation in which first of all they performed the preprocessing steps such as image enhancement and thinning operation then postprocessing steps that is all feature extraction stage for pattern matching. They also use the MD5 hash function in the final result for security purpose. According to them, using these techniques the performance is improved in terms of efficiency as well as speed also [1].

Pokhriyal et al. implemented a method called MERIT (Minutiae Extraction using Rotation Invariant Thinning). In this method a fingerprint image is thinned irrespective of the fingerprint's position and then extracts minutiae points of fingerprint image. First, binarization process is performed in which the fingerprint image is converted into binaries and then 0–1 pattern. After some morphological operations are performed such as dilation and erosion and some if then rules for a 3×3 mask, that is, to be convoluted on the image to skeletonize on applying.

Finally, some postprocessing is done to remove false minutiae structures on the thinned fingerprint image [2].

Al-Ani et al. proposed to develop a novel system, an algorithms which improves the performance of an existing fingerprint algorithm. In this algorithm, they concentrate on the thinning process and optimum result is adequate when related to other system according to them [3].

Fernando Alonso-Fernandez, Julian Fierrez-Aguilar, Javier Ortega-Garcia in 'A Review of Schemes for Fingerprint Image Quality Computation' describes the importance of FRS to improve the quality of fingerprint image. They also use the MCYT database for implementation, testing, and comparing a selection of them including 9000 fingerprint images and almost all algorithms behave similarly in results [4].

According to '**Om Preeti Chaurasia**,' if a good quality input fingerprint image is processed in a particular order then definitely it will produce the better result in terms of minutiae extraction and vice versa is also true. So we can say that the performance of fingerprint image is fully dependent on the quality of input image. In other words if the good quality input fingerprint image is proceed then obtained result is also good for fingerprint matching based on minutiae points. So the output of fingerprint image is always equal to the input fingerprint image in terms of performance or quality of image [5].

'The Impact Performance of the user attempts on FRS' by Dr. NeerajBhargava, Dr. RituBhargava, Manish Mathuria, Minaxi Cotia.' Result of the experiment tells us the affected performance of the Minutia matching-based FRS, through the false acceptance rate (FAR) and false rejection rate (FRR) [6].

In 'Fingerprint Minutia Match Using Bifurcation Technique,' Ravi Kumar proposed a simple technique to improve the quality of input image in such a way that minutiae extraction is performed in a comfortable way using the combination of several techniques of image preprocessing including the two categories of minutia and bifurcation [7].

3 Role of Quality Factors

The identification of accurate and reliable fingerprint image relies on the quality of fingerprint images. It highly affects by the performance of fingerprint recognition system. The FRS is very sensitive to the image quality degradation or to the noise. If the quality of fingerprint images is better than the performance of system in terms of 'Minutia Extractions' also gives the better results. If the poor fingerprint quality images are used then result also is spurious (Tables 1 and 2).

In many applications areas, it is preferable to estimate the quality of captured fingerprint images rather than to precede the enhancement of input image.

To achieve the better performance from higher quality images, the poor quality fingerprint images or degraded images can be eliminated or to replace these images. The factors that evaluate the quality of fingerprint image are as follows:

Table 1 Environmental factors and their states affecting FRS [8]

Factor		State
Environmental humidity		0–100 %
Environment temp (C)	<0	Winter
	0–10	Starting of the spring or end of the fall
	10–20	Spring or fall
	20–30	At room temperature
	>30	Summer
User pressure	High	Strong pressure
	Middle	Normal pressure
	Low	Soft pressure
Skin humidity	High	71–100 %
	Middle	36–70 %
	Low	0–35 %

Table 2 Computation points

S. no.	Filter	Ridge point	Bifurcation point
1	Input image	580	2074
2	'Unsharp'0.5	1624	2953
3	'Gaussian'5	349	1479
4	'Average,'2	197	1360
5	'Disk,' 0.5	580	2074

3.1 Device Conditions

The quality of fingerprint images can also influence the acquisition device conditions which are dirtiness, noise, sensor, size, and time etc. It contains a sensor which is characterized by area, dpi, and dynamic range.

- Contrast and the level of distortions are effects of the pressure of the finger on to the device and the distribution of pressure in the area of contact.
- Finger pattern will be changed from readout to readout if the position of it is unstable.

3.2 Environmental Factors and Skin Conditions

Skin conditions such as, skin moisture, air temperature, air humidity, dryness, wetness, dirtiness, temporary or permanent cuts, and bruises, etc., and environmental factor can influence the image quality. For example a dry finger contains the white area and breaks the line definition but in other case the wet finger reflects the black areas without line definitions.

In some cases when the coverage are of fingerprint is too narrow then the data do not show up. In some conditions of skin like scars, creases, wrinkles are inherent and cannot be replaced. But in other cases there are no permanent level or user behavior can easily be adjustable as per requirements.

4 Methodology

The given fingerprint images are poor quality fingerprint images. So quality of fingerprint images is improved by applying the some image processing steps and using the unsharp filter that gives the better result in comparison to some other filters.

INPUT: Two similar fingerprint images are acquired.

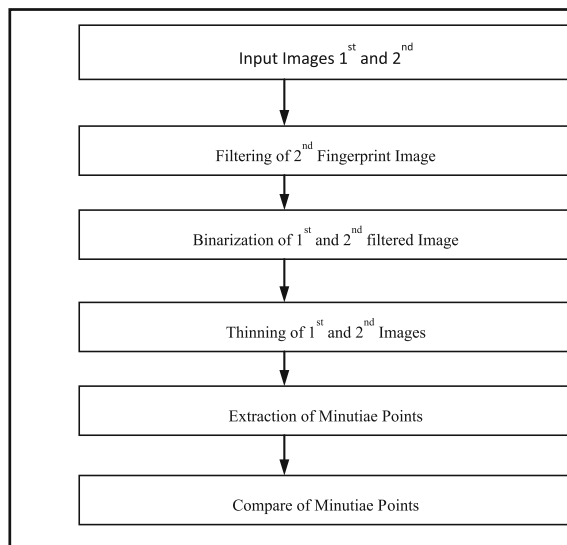
OUTPUT: Improved quality of given fingerprint image

Step 1: Two similar fingerprint images are acquired, i.e., first fingerprint image and second fingerprint image.

Step 2: Enhancement of the first fingerprint image and second fingerprint image using some image preprocessing steps such as

- (a) Filtered the second fingerprint image.
- (b) Binaries the first fingerprint image and second filtered fingerprint image.
- (c) Thinning of first fingerprint image and second filtered fingerprint image is performed using some morphological process.

Step 3: Minutia points are extracted from first and second filtered fingerprint images.



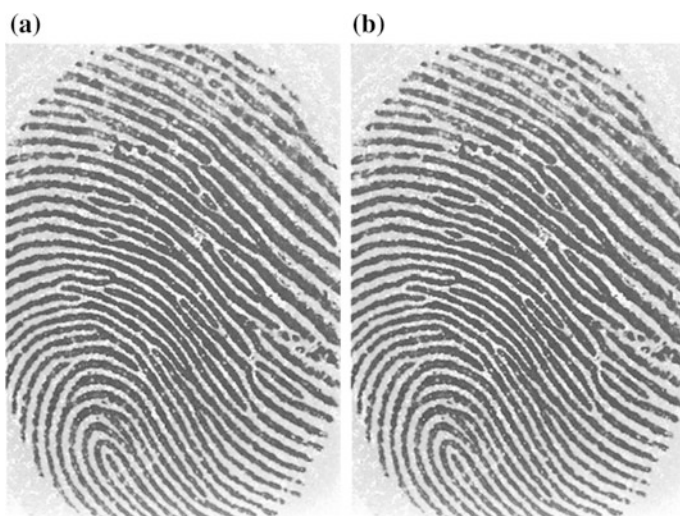


Fig. 1 Input image first and input image second

Flow chart of FRS

Step 4: Counts the minutia points of first and second filtered fingerprint images and finds out that how much fingerprint image quality is improved.

Enrollment/Image Acquisition In this phase, two similar fingerprint images are captured from the database. The acquired images may be noisy which can reduce the quality of a fingerprint image. Therefore the performance of FRS can also be affected. So for robust the quality of input fingerprint image filtering is applied on second fingerprint image [9] (Figs. 1, 2, 3, 4 and 5).

4.1 Filtering

Binarization of Image The process of converting a gray scale image into black and white image is called as binarization. In which the black lines in image is called as ridges where as the white area between the ridges is called as valleys. Threshold process is used to execute the modification of the gray scale image into binary image. Finally gray scale image fully convert into binary image with each pixel value either 0 or 1. After this modification the fingerprints valleys are spotted with white color while ridges are spotted with black color. In MATLAB “`imb2w`” function performs binarization.

Thinning After binarization process the next preprocessing technique used for quality improvement is thinning. The thinning process is used for destructing the redundant pixels and decreases the ridges lines into single pixel wide.

Filtering:

Fig. 2 Filtering of second input image



Fig. 3 Binarization of first and second image

The running process is done using the MATLAB's inbuilt thinning function, that is, 'bwmorph.' Example-bwmorph ('Binary image,' 'thin,' Info);

Minutiae Extraction After the preprocessing steps of fingerprint mage, minutia extraction phase is carried out. Based on the extracted minutiae points quality improvement of filtered image is computed.

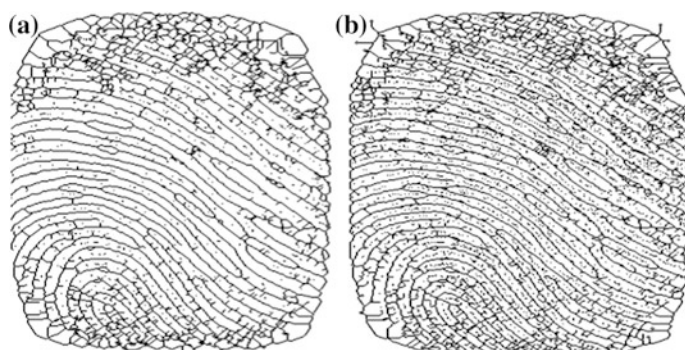


Fig. 4 Thinning of first and second image

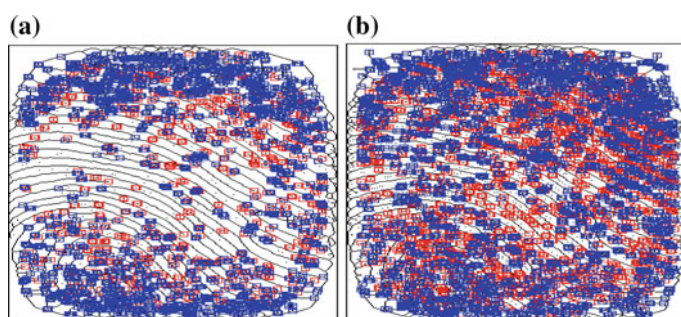


Fig. 5 Minutia extraction of first and second image

5 Result

The experimental results exposed that using image preprocessing steps and filtering function (unsharp filter) the quality of fingerprint image is improved.

6 Conclusion

The research on FRS provided knowledge about new developments required for quality improvements. There are some specific conditions which affect quality of FRS, sometimes fake access. This research work is accomplished by implementing an algorithm to read fingerprint image and to calculate total number of ridges and bifurcation points. Fingerprint recognition is based on minutiae matching which is only possible by extracting minutiae information from the fingerprint image. The quality of fingerprint image affects the information extracted from the image so the quality of FRS.

The quality is affected by either human error or machine capturing quality and sometimes environmental factors. To improve the quality of FRS some special image processing operations are used such as filtering. The result varies according to the image quality and affected area. There some filters are applied and finally scores are calculated that show the possibility of quality improvement of FRS using filtering operation.

References

1. Ali H. A. and Ne'ma B. M., "Multi-Purpose Code Generation Using Fingerprint Images", 6th WSEAS International Conference on Information Security and Privacy, Tenerife, Spain, December 14–16, 2007.
2. Avinash Pokhriyal et. al., "MERIT: Minutiae Extraction using Rotation Invariant Thinning", International Journal of Engineering Science and Technology Vol. 2(7), 2010, 3225–3235.
3. Muzhir Shaban Al-Ani et. al., "An Improved Proposed Approach for Hand Written Arabic Signature Recognition", Advances in Computer Science and Engineering Volume 7, Number 1, 2011, Pages 25–35.
4. Fernando Alonso-Fernandez and (in alphabetical order) Josef Bigun, Julian Fierrez, Hartwig Fronthaler, Klaus Kollreider, Javier Ortega-Garcia, "A Review of Schemes for Fingerprint Image Quality Computation".
5. Om Preeti Chaurasia, "An Approach to Fingerprint Image Pre-Processing", I.J. Image, Graphics and Signal Processing, 2012, 6, 29–35.
6. Dr. NeerajBhargava, Dr. RituBhargava, Manish Mathuria, Minaxi Cotia, "Performance Impact of the user attempts on Fingerprint Recognition System (FRS)", International Journal Of Emerging Technologies and Applications In Engineering Technologies And Sciences (IJ-ETA-ETS).
7. L. Ravi Kumar, S. Sai Kumar, J. Rajendra Prasad, B. V. Subba Rao, P. Ravi Prakash " Fingerprint Minutia Match Using Bifurcation Technique", S Sai Kumar et al, International Journal of Computer Science & Communication Networks, Vol 2(4), 478–486, Sep. 2012.
8. Shan Juan Xie, JuCheng Yang, Dong Sun Park, Sook Yoon and Jinwook Shin " Fingerprint Quality Analysis and Estimation for Fingerprint Matching" www.intechopen.com.
9. <http://bias.csr.unibo.it/fvc2000/databases.asp>.

A Composite Approach to Digital Video Watermarking

Shaila Agrawal, Yash Gupta and Aruna Chakraborty

Abstract Digital media stands today as one of the most effective forms of communication. The extensive use and popularization of this form of media is attributed to the ease of generation and distribution. Thus, the protection of copyrights of digital data is a prerequisite in the distribution. This paper proposes a composite technique of digital video watermarking using discrete wavelet transformation in the odd-numbered video frames, with a discrete cosine transformed watermark and singular value decomposition in the even-numbered ones with the purpose of extracting a composite watermark by superimposing the individual extracted watermarks which shall be more similar to the original watermark under most attacks. The result of the proposed algorithm shows a maximum value of PSNR equal to 65 dB and minimum value of 45 dB under sharpening filter.

Keywords Discrete wavelet transform · Discrete cosine transform · Singular value decomposition · Digital video watermarking

1 Introduction

Digital media is one of the most widely used means of communication in the present-day. The authentication of the same is an important objective associated with their distribution and use. Any data that demands a high security level needs sound

Shaila Agrawal (✉)

Information Technology Department, St. Thomas' College of Engineering and Technology, Kolkata, India
e-mail: shaila.agrawal@gmail.com

Yash Gupta · Aruna Chakraborty

Computer Science and Engineering Department,
St. Thomas' College of Engineering and Technology, Kolkata, India
e-mail: yashino95@gmail.com

Aruna Chakraborty

e-mail: aruna.stcet@gmail.com

copyright protection and identification. Digital watermarking [1, 2] is a technology used to preserve the copyrights of the digital media while retaining other characteristics of it. A watermark is a logo or an authoritative information of the owner. The watermark can be extracted in case of any dispute pertaining to ownership.

Video watermarking is similar to image watermarking where the video can be divided into a number of frames with all the frames being watermarked individually [3, 4]. The work on compressed videos is given in [5–12]. Due to high frame rate of videos, they are more susceptible to attacks such as frame averaging, frame dropping, and frame swapping [13]. Statistical analysis can also be used to detect the watermark, making it easier to access and to make changes in the watermark. The embedding process of the watermark can be carried out in two different ways, namely, spatial domain watermarking and transform domain watermarking [14]. In the former, the pixel intensity values are directly modified according to the watermark, whereas in the latter, the pixels are chosen according to an algorithm and only the intensity values of those pixels are modified. Thus, transform domain watermarking schemes ensure more imperceptibility and randomness in the distribution of the watermark and have also proven to enhance robustness against geometric attacks. Some recent research works use transform-domain techniques like discrete wavelet transform (DWT) [15, 16] discrete cosine transform (DCT) [17], singular value decomposition (SVD) [18–20], principal component analysis (PCA), and discrete fourier transform (DFT).

The present-day convention uses the same watermarking technique for all the frames. In this paper, we have proposed a novel approach where we have used a composite scheme of applying DWT in one half of the frames (odd-numbered ones) with a DCT transformed watermark and of applying SVD in the other (even-numbered ones). DWT with a DCT watermark produces a watermarked video that is more robust to geometrical attacks and SVD yields a more similar extracted watermark to the original. The novelty of the paper, lies in using two distinct approaches, i.e., is to make the watermark more robust to different attacks while attenuating the quotient of similarity between the extracted and the original watermarks.

The paper is organized as follows. Section 2 discusses the watermarking schemes in details. Section 3 contains the algorithms and the proposed method. Section 4 deals with the experiment and result analysis, and Sect. 5 provides the conclusions and scope.

2 Watermarking Scheme

The watermarking method which has been used in the paper utilizes three basic techniques: DWT, DCT, and SVD. As mentioned earlier, we shall apply two different techniques of embedding a watermark, one to the even-numbered frames and the other to the odd-numbered frames. To the odd-numbered frames, DWT has been applied with a DCT watermark while to the even-numbered frames, we have applied SVD.

2.1 Discrete Wavelet Transform

Discrete wavelet transform (DWT) is used widely in the field of signal processing. A two-dimensional DWT is a combination of two single 1-D DWT's applied to both the horizontal and vertical directions. 2-D DWT is used to decompose the image into lower resolution approximation sub-band (LL) as well as horizontal (HL), vertical (LH), and diagonal (HH) detail components, (Fig. 1) [13]. The embedding process is carried out in the LL sub-band to make the watermarked image withstand lossy compression.

2.2 Discrete Cosine Transform

Discrete cosine transform (DCT) is another commonly used transform technique in signal processing. It aims at converting an image from its spatial domain to frequency domain in order to make it robust against different attacks like contrast adjustment, low pass filtering, etc. discrete cosine transform is defined by the Eqs. (1) and (2).

$$f(mn) = a(j)a(k) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(mn) \cos \left[\frac{(2m+1)j\pi}{2N} \right] \cos \left[\frac{(2n+1)k\pi}{2N} \right] \quad (1)$$

The inverse discrete cosine transform is given as:

$$f(mn) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a(j)a(k)f(mn) \cos \left[\frac{(2m+1)j\pi}{2N} \right] \cos \left[\frac{(2n+1)k\pi}{2N} \right] \quad (2)$$

LL1	HL_1	LL3	HL3	HL2	HL1
		LH3	HH3		
LH_1	HH_1	LH2		HH2	
		LH1		HH1	

Fig. 1 DWT sub-bands

2.3 Singular Value Decomposition

Singular value decomposition is a technique for expressing a matrix as a product of a diagonal matrix and two orthogonal matrices. The SVD of a given image A in the form of a matrix is defined by Eq. (3).

$$A = USV^T \quad (3)$$

where U and $V \in \mathbb{R}$ are $N \times N$ unitary matrices and $S \in \mathbb{R}$ with dimensions $N \times N$ is a diagonal matrix (Fig. 2).

3 Proposed Watermarking Algorithms

The algorithms proposed for embedding and extracting the watermark are given below.

3.1 Watermark Embedding

- Step 1: Input the original video and extract the individual frames.
- Step 2: For odd-numbered frames follow Algorithm 1, cited by M. Chaturvedi and Dr B.J. Basha [16].
- Step 3: For even-numbered frames follow Algorithm 2, based on SVD [17, 19].
- Step 4: Combine the outputs obtained from step 5 of Algorithm 1 and from step 6 of Algorithm 2 to obtain the watermarked video.

3.1.1 Algorithm 1

- Step 1: Input the watermark, change it to YUV form (gray scale) from RGB form and extract the luminance or the Y component of it.
- Step 2: Apply 2-D DCT to the extracted Y component.

Fig. 2 Representation of Singular Value Decomposition

$$SVD(A) = \begin{bmatrix} U_{1,1} & . & . & U_{1,n} \\ U_{1,1} & . & . & U_{2,n} \\ . & . & . & . \\ U_{n,1} & . & . & U_{n,n} \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & \sigma_m \end{bmatrix} \begin{bmatrix} V_{1,1} & . & . & V_{1,n} \\ V_{2,1} & . & . & V_{2,n} \\ . & . & . & . \\ V_{n,1} & . & . & V_{n,n} \end{bmatrix}^T$$

- Step 3: For every odd numbered frame, change it from RGB form to YUV form and apply 2-D DWT to the Y component of it.
- Step 4: Resize the 2-D DCT transformed watermark according to the LL part of the image and embed it into the LL part with a strength alpha by Eq. (4)

$$LL1 = LL + \alpha * WM \quad (4)$$

Here, LL is the approximation image obtained from 2-D DWT, WM is the resized watermark matrix and LL1 is the watermarked approximation image.

- Step 5: Reconstruct the original frame from LL1 by applying inverse 2-D DWT to the Y component and changing it back to RGB form.

3.1.2 Algorithm 2

- Step 1: Input the watermark, change it from RGB to YUV form and apply SVD on the Y component of it.
- Step 2: To the even numbered frames extracted from step 1 of Algorithm 1, change the color format from RGB to YUV and extract the Y component of it.
- Step 3: Apply SVD on the Y component of the frames.
- Step 4: Embed the watermark into the frames with strength alpha (same as that in Algorithm 1), using Eq. (5).

$$Sf1 = Sf + \alpha * Sw \quad (5)$$

Here, Sf is the singular matrix obtained from single value decomposition of the video frame, Sw is the singular matrix obtained through single value decomposition of the watermark and Sf1 is the final singular matrix.

- Step 5: Reconstruct the Y component of the watermarked frame by applying Eq. (6).

$$Y1 = Uf * Sf1 * Vf' \quad (6)$$

where Uf and Vf are the matrices obtained by SVD of the video frame.

- Step 6: Change the watermarked frame obtained from YUV to RGB format.

3.2 Watermark Extraction

- Step 1: Input the watermarked video, together with the matrices, U_w and V_w obtained by performing SVD on the watermark image.
- Step 2: Extract the individual frames. For odd-numbered frames, follow Algorithm 1 for extraction, cited by M. Chaturvedi and B. J. Basha in [16].
- Step 3: For the even-numbered frames, follow Algorithm 2, cited in [18, 20].
- Step 4: Resize the image matrices obtained from step 5 of Algorithm 1 and from Step 5 of Algorithm 2 to match their dimensions.
- Step 5: Add the two images to obtain the final watermark image.

3.2.1 Algorithm 1

- Step 1: Convert the frames from RGB TO YUV color format and separate the Y component.
- Step 2: Perform 2-D DWT on the Y component.
- Step 3: From the LL1 component obtained from step 3, obtain the watermark by the following equation,

$$WM = (LL1 - LL)/\alpha \quad (7)$$

Here, WM is the embedded watermark component and LL is the Y component of the original video frame.

- Step 4: Perform inverse 2-D DCT on the output of step 4 to obtain the Y component of the watermark.
- Step 5: Convert the YUV watermark image to RGB format.

3.2.2 Algorithm 2

- Step 1: From the extracted even-numbered frames, convert the color format of the image from RGB to YUV.
- Step 2: Obtain the Y component and perform SVD on it to get U_{f1} , S_{f1} and V_{f1} .
- Step 3: From S_{f1} , obtain S_w using the equation,

$$S_w = (S_{f1} - S_f)/\alpha \quad (8)$$

- Step 4: From S_w obtained from the previous step, reconstruct the Y component using the following equation,

$$Y = U_w * S_w * V_w' \quad (9)$$

where U_w and the V_w are the matrices obtained from the SVD of the original watermark in the embedding process.

Step 5: Convert the image from YUV to RGB color format.

4 Experimental Results

The algorithm given in Sect. 3 is tested on a video named ‘Demo.wmv’ and a watermark image named ‘Sample.jpg’. Since, the first algorithm embeds a watermark of the size of the LL sub-band and the second embeds that of the size of the frame, the extracted watermarks are resized to 320×320 images and then superimposed to produce the final extracted watermark. The calculations are done by resizing the original watermark image to 320×320 as well.

The performance of the algorithm is measured in terms of the imperceptibility and robustness against possible attacks like filtering, noise addition, geometric attacks, etc.

4.1 Peak Signal to Noise Ratio

The peak signal to noise ratio (PSNR) is a measure of the imperceptibility of the watermark. A high value of PSNR indicates more imperceptibility. The PSNR is calculated by (10).

$$\text{PSNR} = 10 \log_{10}(\text{MAX}_i^2 / \text{MSE}) \quad (10)$$

where MAX_i is the maximum possible value of a pixel in an image, MSE is the mean squared error which is calculated by (11).

$$\text{MSE} = \left(\frac{1}{mn} \right) \sum_{i=0}^m [I(i,j) - I'(i,j)]^2 \quad (11)$$

where I and I' are pixel values at location (i, j) of the original and the extracted frames, respectively.

4.2 Normalized Correlation Coefficient

The normalized coefficient (NC) is used to determine the robustness of the watermarking and has a peak value of 1 [21]. The formula for calculating NC is given in (12).

Fig. 3 Original video frame

$$NC = \frac{\sum_i \sum_j W(i,j)W'(i,j)}{\sqrt{(\sum_i \sum_j W(i,j))\sqrt{(\sum_i \sum_j W'(i,j))}}} \quad (12)$$

where W and W' represent the original and extracted watermarks, respectively.

Algorithm 1 and 2 are applied on the video, 'Demo.wmv' Fig. 3 with the watermark given in Fig. 4 and the extracted watermark is given in Fig. 5.

The value of PSNR and NC for the proposed method is found to be 65.073 dB and 0.547, respectively. These are the values without noise or any other attack. The values of both PSNR and NC under some of the attacks are given in Table 1. The video frames after the addition of salt and pepper attack and sharpening filter are given in Figs. 6 and 7.

Fig. 4 Original watermark

Fig. 5 Extracted superimposed watermark



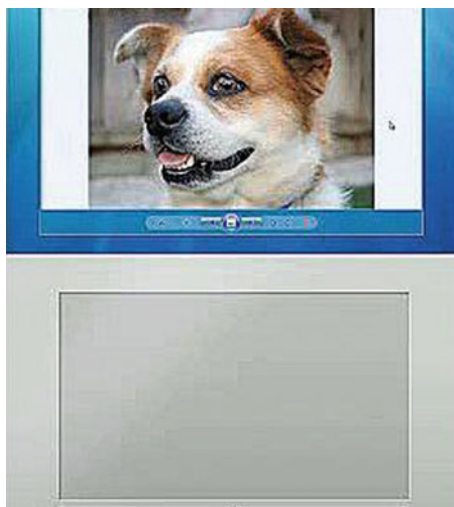
Table 1 Values of PSNR and NC under different attacks

Attacks	PSNR (dB)	NC
Gaussian noise	53.397	0.509
Salt and pepper noise	47.749	0.349
Poisson noise	65.097	0.547
Median filtering	61.356	0.530
Sharpening filter	45.532	0.968

Fig. 6 Video frame after the addition of salt and pepper



Fig. 7 Video frame after using sharpening filter



5 Conclusions and Scope

In highly secure digital media, it is as important to have a good assertive proof as it is to make it imperceptible. The proposed method satisfies both the criteria to a good extent and is robust to more number of attacks than the individual techniques used. However, this method is more complex to implement than the conventional ones. Thus, it is more convenient to be implemented in applications that require high security and authentication. This method can also be modified to use three or more schemes to produce more robust techniques by dividing the video frames into three or more such different sets and by applying three different schemes to them. This paper uses modulo 2 function to generate two different sets. The code can be further randomized to yield better results and more robust watermarking procedures.

References

1. M. Natarajan, G. Makhdimil: Safeguarding the Digital Contents, *Digital Watermarking: DESIDOC Journal of Library & Information Technology*, vol 29, pp. 29–35 (2009).
2. C.I. Podilchuk, E.J. Delp: Digital Watermarking Algorithms and Applications: *Signal Processing Magazine*, vol 18, pp 33–46, IEEE (2001).
3. Yeo, M.M. Young: Analysis and Synthesis for new Digital Video Applications: *International Conference on Image Processing*, vol 1, pp. 1 (1997).
4. G. Doerr, J.L. Dugelay: Security Pitfalls of Frame-by-Frame Approaches to Video Watermarking: *Signal Processing, IEEE Transactions*, vol 52, pp. 2955–2964 (2004).

5. M.K. Thakur, V. Saxena, J.P. Gupta: A Performance analysis of Objective Video Quality Metrics for Digital Video Watermarking, 3rd IEEE International Conference, vol 4, pp. 12–17 (2010).
6. Voloshynovskiy, S. Pereira, T. Pun: Watermark attacks: Erlangen Watermarking Workshop (1999).
7. G. Langelaar, I. Satyawan, and R. Lagendijk: Watermarking Digital image and Video Signal Processing, vol, pp. 20–46 (2000).
8. F. Hartung and B. Girod: Watermarking of uncompressed and compressed video: Signal Processing, vol 68, pp 283–301 (1998).
9. T. Khatib, A. Haj, L. Rajab, H. Mohammad: A Robust Video Marking Algorithm: Journal of Computer Science, vol 4, pp 910–915 (2008).
10. J. Meng, S. Chang: Embedding Visible Video Watermarks in Compressed Domain, International Conference on Image Processing, Proceeding vol 4, pp 474–477 (1998).
11. G. Doerr, J.L. Dugelay: A guide tour of video watermarking: Signal Processing: Image Communication, vol 18, pp 263–282 (2003).
12. Y.R. Lin, H.Y. Huang, W.H. Hsu: An embedded watermark technique in video for copyright protection: IEEE, 18th Conference on Pattern Recognition, vol 00, pp 795–798 (2006).
13. Department of Electronics and Communication, ece.mits.ac.in.
14. Resource Library, www.eurojournals.com.
15. M. Jianshang, L. Sukang, T. Xiaomei: A Digital Watermarking Algorithm based on DCT and DWT: Proceedings of International Symposium on WISA (2009).
16. M. Chaturvedi, Dr. B. J. Basha: Comparison of Digital Image Watermarking methods DWT & DWT-DCT on the basis of PSNR: International Journal of Innovative Research in Science, Engineering and Technology, vol 1, Issue 2 (2012).
17. M. M. Rahman: A DWT, DCT and SVD based watermarking technique to protect the image piracy: International Journal of managing Public sector information and Communication Technologies (IJMPICT), vol 4, no. 2 (2013).
18. H. Shaikh, M.I. Khan, Y. Kelkar: A Robust DWT Digital Image Watermarking Technique Basis of Scaling factor: International Journal of Computer Science, Engineering & Applications, vol. 2, no. 4 (2012).
19. Seema, S. Sharma: DWT-SVD Based Efficient Image Watermarking Algorithm to Achieve High Robustness and Perceptual Quality: International Journal of Advance Research in Computer Science and Software Engineering, vol 2, Issue 4 (2012).
20. L. Lai, C. Tsai: Digital Image Watermarking using Discrete Wavelet Transform and Singular Value Decomposition.
21. S. Sinha, P. Bardhan, S. Pramanick, A. Jagatramka, D.K. Koley, A. Chakraborty: Digital Video Watermarking using Discrete Wavelet Transform and Principal Component Analysis, International Journal of Wisdom Based Computing, vol 1 (2011).

Effective Congestion Less Dynamic Source Routing for Data Transmission in MANETs

Sharmishtha Rajawat, Manoj Kuri, Ajay Chaudhary
and Surendra Singh Choudhary

Abstract MANET is widely used for Ad hoc-based communications, but it suffers from several deficiencies like mobility cause change in topology and network partitions. The packet losses and congestion is very common scenario in MANET. In this work we provide a novel routing scheme which is able to provide consistent network stability and reliable packet delivery. We develop a routing scheme based on reactive route discovery process related to dynamic source routing protocol by proposing an optimal least busy path effective congestion less scheme which tries to minimize the packet drop keeping in mind the end goal which is to provide high quality path. The quality of our scheme is least busy routes, link stability, path quality, and high response time. The simulation results demonstrate the effectiveness of our proposed method; it is largely unaffected with increase in network mobility and network load. It has been likewise seen that the packet delivery ratio has been increased and the first packet received with the end-to-end delay has been decreased when our proposed method is compared with the conventional DSR protocol.

Keywords MANET · DSR, routing protocols · Ad hoc networks · Mobility

S. Rajawat (✉) · A. Chaudhary · S.S. Choudhary
Department of CSE & IT, Government Engineering College Bikaner,
Bikaner 334003, India
e-mail: sarmirajawat@gmail.com

A. Chaudhary
e-mail: ajaychaudhary@acm.org

S.S. Choudhary
e-mail: surendra2060@gmail.com

M. Kuri
Department of ECE, Govtment Engineering College Bikaner, Bikaner, India
e-mail: kuri.manoj@gmail.com

1 Introduction

Well-connected collection of routers through wireless medium is known as MANET, which also allows to setup network using mobile nodes in a cost-effective manner without requiring a fixed network topology [1, 2]. Nowadays congestion in networks is very common because the number of people using mobile communication is growing rapidly. Due to the heavy traffic there is a possibility that the radio link may break and it need to repair or replace. These routing operating operations degrades the routing performance. So the route discovery process should be changed in such a way that it can discover routes in terms of other metrics that are more reliable, effective, and suitable other than the lowest hop count. The classification of exiting MANET routing protocols is mainly done relying upon routing strategy and network structure. As indicated by the routing strategy the routing schemes can be classified as proactive and reactive. Whereas, relying upon the network structure these are named flat routing, hierarchical routing, and location aware routing. The proactive, reactive, and hybrid schemes go under the flat routing protocols. Routing schemes that do not keep routing information are known as reactive protocol [3], this looks for the route in an on-demand way. In proactive protocols node in the network keeps up the routing information of other nodes before it is required, hence they are called proactive protocols. Route information is kept in the routing tables and must be updated with topology change [3]. The hybrid protocols adopt features from both reactive and proactive protocols.

1.1 Our Contributions

This research work makes the following contributions: To discover the optimal route we propose an effective congestion less routing scheme which adopts network reliability and stability as routing criteria. We are introducing a new metric to measure reliability achieved in dynamic source routing (DSR) protocol [4]. To make sure that the discovered routes are of good quality in terms of less traffic load we make sure that it contains less queue load in comparison of other existing links which makes the impact of congestion less significant leading to reduced risk of data packet being dropped by received signal that implies, one gets assurance that link has lower traffic with higher response time and higher stability as compared to other alternative links.

2 Proposed Approach

In MANETs the routing protocols find routes based on shortest hop count metric [5]. Route selected in this way is not always the best path as there is the possibility that the selected route is less reliable in terms of stability, congestion, response

time, and high interference if the network is loaded with heavy traffic and high mobility. So, by keeping this in mind a new routing protocol was developed that uses the queue load information of the received RREQ message. The question was whether to include the current link or not in the route, or select the lowest loaded link among the all existing links. The value of the queue load of RREQ is calculated using its received hop count value during the route discovery phase. The functionality of our proposed design is having three phases:

2.1 Route Discovery (Route Request Phase)

This phase tries to find route in such a way that the probability of their breakage is lower during the long communication. To achieve this we use a queue load-based route discovery process as described in the algorithm and flowchart. When a source gets a packet to send it checks its entry in cache whether it is valid or not. If it has valid entry then the data packet is transmitted to the next hop toward destination using the route in cache. If it does not have a valid entry it executes it to the route records of the RREQ and hence starts to broadcast the RREQ packet. This RREQ propagates to reach either to destination or an intermediate hop with route to the destination. The RREQ contains: source node address, destination address, and hop count value (Fig. 1).

Our proposed EC-DSR protocol uses two additional data structures. First is the result of slight modification of RREQ control message as QL_INFO field is added by physical layer upon receiving the RREQ signal. This field contains information about network load which is mainly used by network layer. The information field QL_INFO is added with RREQ to pass the value of queue load. Queue load value of a received message is calculated using its received hop count value which is extracted from RREQ message format. The hop count is calculated as a number of received RREQ signals. Queue load information is further used by network layer in order to select the best route in terms of lowest load as compared to all routes between source destination pair. RREQ_BUFFER_cache created at each node serves as the second data structure. The buffer comprises of some fields as follows:

- (i) Source address: it gives the source address of the message received.
- (ii) Flooding_Id: it gives the flooding of the received message (RREQ).
- (iii) RREQ_message: it gives the received message (RREQ). As soon as the RREQ is received the hop buffers in the given buffer, which is implemented by us at each

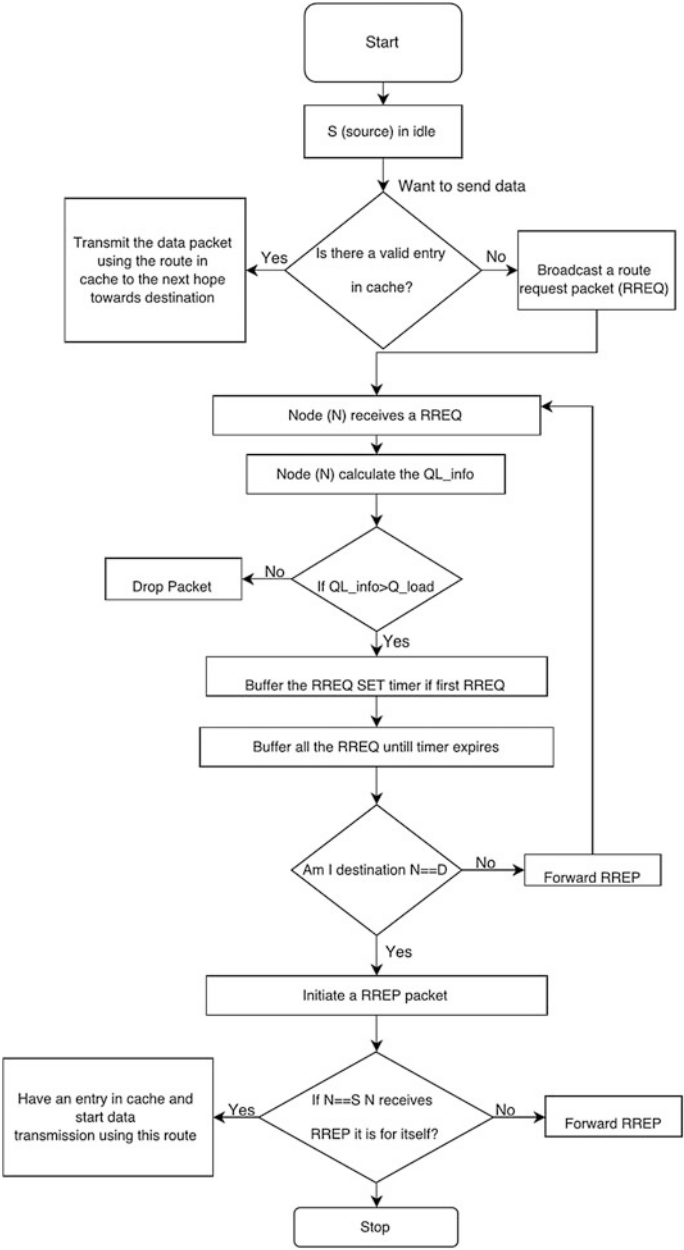


Fig. 1 Flow chart of EC-DSR routing protocol

hop, is available in the network. The node starts a timer of some 15 μ s when the RREQ message is inserted in the buffer. All the duplicate messages received from other neighboring nodes are also buffered by this node until this timer expires. The RREQ message kept in the buffer along with the messages having the lowest load value is kept at the top of the message buffer that are sorted by the node, when the timer is associated with message storing buffer. After sorting the RREQ, the node extracts the top RREQ and processes it. The remaining messages in the RREQ buffer are discarded by the node. Until the RREQ reaches its destination, the process is applied to all the intermediate nodes.

2.2 Route Reply Phase

First phase is completed when destination node receives the RREQ signal. Until the timer expires, the destination node stores the received RREQ in the buffer. As the timer expires, the top RREQ is extracted by destination hop from the buffer. The node through which the message is received is marked as the next hop towards the source of the message received. Further the destination node uses the next hop node to send the RREP message to the source node. After this a unicast ack is sent back to the source node. The generated RREP message is sent by the destination node following the same reverse route created during the RREQ propagation phase.

2.3 Route Maintenance

As the link between a hop and the next node on the route fails the maintenance process is begun. Node failed in packet delivery will send an error message (control message) to source and then all broken link nodes are removed from cache. Then either route discovery is performed, else different cached path is selected. Error message is shared to each node which has transmitted the packet using broken route. This maintenance process is started due to node mobility. Best forwarding service is provided by DSR protocols which can be enhanced by augmenting it with QoS functionality to provide high priority preferential treatment.

Definations	
<i>S</i> - Source node	<i>Destination node</i> - Destination node
<i>Int_node</i> - Intermediate node	<i>R_buf</i> - RREQ message buffer
<i>ql_load</i> - Load at queue 1	<i>RT</i> - Routing table of anode
<i>Tx</i> - Timer of <i>R_buf</i> at node <i>x</i>	

Algorithm

```

if S got data packet for D then
  if S not have route for D in its RT then
    S starts the EC-DSR protocols route discovery process
    if Receive a fresh RREQ or duplicate message then
      Int_node increments the Hop count in the packet by one forwards it
      to all the nodes and calculates the ql_load and add it in the
      QL_INFO field of RREQ message;
      Int_node store the RREQ message in its R_buf and set the timer if
      its fresh RREQ;
      When Ti expire node Int_node extract the RREQ with the lowest
      ql_load;
      Node Int_node rebroadcast the extracted RREQ and discard the
      R_buf;
    end
  if D receives the RREQ then
    Int_node store the RREQ message in its R_Buf and set the timer if
    its fresh RREQ;
    When Ti expire node Int_node extract the RREQ with the lowest
    ql_load;
    D gets the pervious hop address of the extracted RREQ message;
    D creates a RREP message and sends it towards S using the
    previous hop selected in last step;
  end
  if S receives the RREP message S updates its RT and sends the buffered
  data packet to D;
end
else
  S send packet to next-hop towards destination node D;
  set Hop count set to zero;
end
end
end

```

Algorithm 1: Route Discovery Process of Proposed

3 Experimental Results

The section discusses the simulations results and comparative investigation of the performance of DSR and proposed EC-DSR protocol on different scenarios over MANETs. EXATA [6] is the simulator using which we have analyzed the feasibility of EC-DSR protocol. To gauge the effectiveness in different kinds of scenarios, outcomes are created on large number of network combinations with different parameters on three different network layer metrics.

3.1 Simulation Model

For characterizing the mobility patterns of individual nodes or entire network the mobility model with particular parameters is determined beneath. As a part of our simulation process we utilized the well know random way point mobility model. Before moving to the target point the arbitrarily picked node's speed is between 0 and 25 m/s and pause time is set to 10 s. Simulations are run for 50 nodes in an area of 1200 m × 1200 m. The 802.11a/g MAC specification is selected for the network and the power selected for transmission is 250 m and is calculated utilizing the nodes transmitting power. The source and destination pairs are selected randomly [7]. The modeling of the source nodes as a data generating node is performed by designing every source node in the network considering CBR (constant bit rate) as traffic generator. It produces the data agreeing; the accompanying specified parameters:

(a) Inter-packet time: 33 ms. (b) Packet size: 512 bytes. (c) Intervals: the beginning and halting time of the CBR sessions.

All the data packets received are stored into the output buffer by every node, while it waits for a path for the destination node. Until the packets are extracted from the buffer by MAC layer for the transmission to physical layer, all the sent packets by the routing layer are put away in the packet queue which is further actualized as a buffer. The data packets are given lower priority in the buffer than the routing packets. Each of the simulations done in this research keep running for certain time loop equivalent to 500 simulated sec. Every data point demonstrated in the charts and tables are speak to a average three keeps running with comparative traffic models, however distinctive arbitrarily created mobility scenario by utilizing diverse seed values.

3.2 Performance Metrics

The performance metrics using which the performance of our work is analyzed.

- *Packet delivery ratio (PDR)*. It is the ratio of the data packets which are sent to the destination nodes with no lapse(S) and the aggregate data packets which are received at destination node(R). The greater value of PDR means the better performance of the protocol. The PDR can be characterized as [2]:

$$PDR = \frac{R(\text{Total number of packets received successfully})}{S(\text{Total number of packet sent from source node})}$$

- *Average end-to-end delay of data packets*. The sent timestamps and the received timestamps at destination are used to calculate the delay. At the end of the simulation to get EED the total number of packets received at destination is divided by the total number of received data packets. The lower value of EED

means the better performance. The average EED is characterized by formula given below [2].

$$\text{EED} = \frac{\text{Delay of each packet received successfully}}{\text{Total number of packets received}}$$

- *First Packet Received.* The time at which first packet is received at the destination node. This implies that the total amount of time to send the file is the initial hand-shake. The destination node uses the packet it receives at first timestamp to compute the delay of all received data packets [8].

3.3 Simulation Results

The results are discussed in a simulated environment using random waypoint (RW) mobility model for diverse parameters considering UDP as the transport protocol with CBR as traffic generator. The feasibility of the EC-DSR is judged against network load and mobility by varied the load and mobility [9]. The evaluation is done by means of simulations using EXATA simulator. It is considered, all the nodes willing to establish connection with other nodes play a major role in the formation of network and the communication among the nodes is within the ad hoc network. Specifically, every node included in the network formation also needs to wish to transmit packets for rest of the other nodes within the network (Table 1).

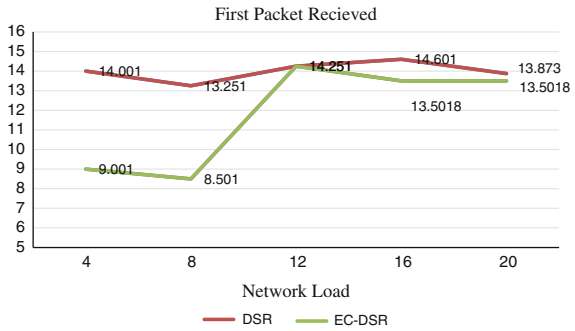
3.3.1 Scenario 1: Varying the Network Load

First packet received. From the Fig. 2 it is seen that the FPR of the EC-DSR is not as high as the DSR because it selects routes with lower queue load which implies

Table 1 Simulation parameters for scenario 1

Parameter	Value
Network load	4, 8, 12, 16, 20
Terrain size	1200 m × 1200 m
Packet size	512 bytes
Traffic type	CBR
Data rate	6 mbps
Performance metrics	First packet received, end-to-end delay, packet delivery ratio
Routing protocol	DSR and EC-DSR
No of nodes	50
Pause time	30 s
Mobility speed	0 to 10 m/s

Fig. 2 FPR versus network load



that the selected route is least busy as compared to the links that are selected if DSR protocol is used. Likewise, the FPR in the DSR protocol increases with increment in the load since load is directly proportional to congestion in the network with this these two things happening firstly, it increases the number of route failures. Secondly, time taken to transmit a data packet to next hop could increment due to the number of re-transmissions.

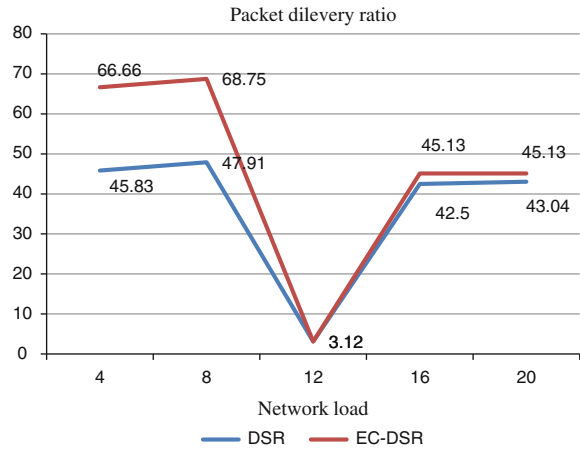
End-To-End Delay. The Fig. 3 depicts that as the load increases, the EED [10] of the data sessions are fluctuating at some points, the general trade shows the EED is directly proportional to network load. Since the load increases, congestion also increases with this these two things occurring. Firstly, it increases the number of route failures. Secondly, the time taken to transmit data packet to next hop could increment due to the increase in the number of re-transmissions. It is observed that EED of EC-DSR is lower at low and moderate network load due to the better selection of routes.

Packet Delivery Ratio. The Fig. 4 depicts that PDR decreases when the network load is increased, if we increase the load the congestion increases resulting in the active routes break and data packets on these routes are dropped. Although, PDR of both comparing protocols remains almost constant in case of lower load whereas it

Fig. 3 Average EED versus network load



Fig. 4 Average PDR versus network load



decreases for both protocols irrespective of the high load, the PDR of EC-DSR is high even in high load networks because it avoids the use of highly loaded links.

3.3.2 Scenario 2: Varying the Network Mobility

First packet received. It is analyzed from Fig. 5 that the FPR increases with the increment in the network mobility. The FPR of EC-DSR is lower than the DSR protocol due to the lower number of re-routing processes because of the selection of high quality route selection process that is implemented in suggested routing protocol (Table 2).

Fig. 5 FPR versus network mobility

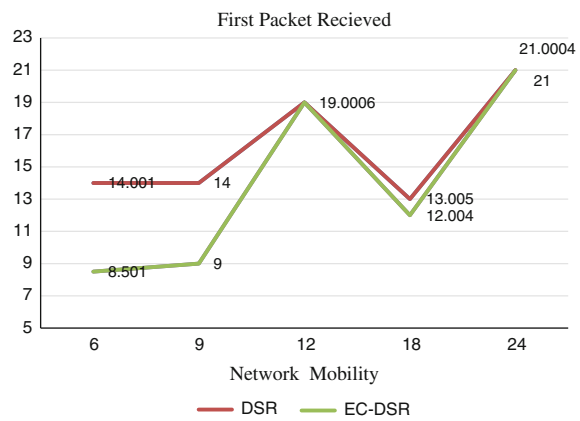
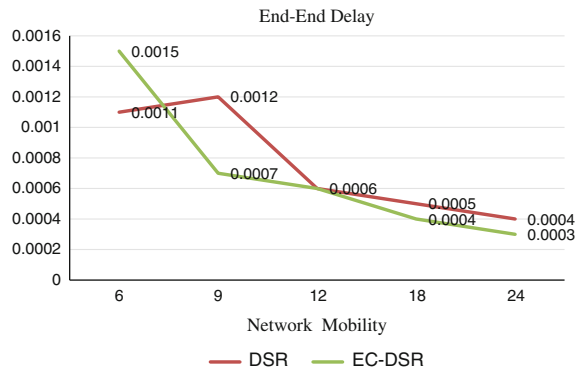


Table 2 Simulation parameters for scenario 2

Parameter	Value
Network mobility	6, 9, 12, 18, 24
Terrain size	1200 m × 1200 m
Packet size	512 bytes
Traffic type	CBR
Network load	8
Performance metrics	First packet received, end-to-end delay and packet delivery ratio
Routing protocol	DSR and EC-DSR
No of nodes	50
Pause time	30 s
Mobility speed	0–10 m/s

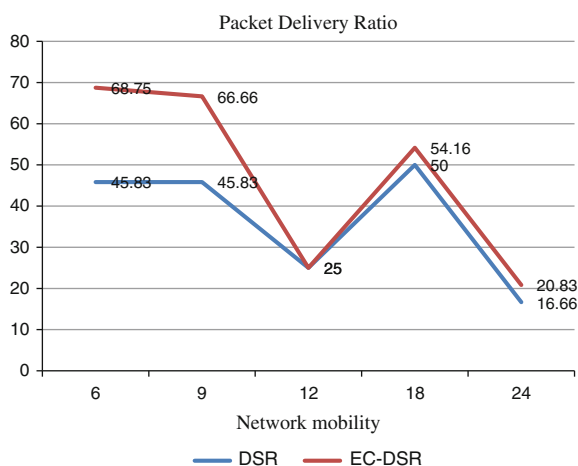
Fig. 6 Average EED versus network mobility



End-To-End Delay. It can be observed from the Fig. 6 that as the mobility increases, the EED of the data sessions are fluctuating at some mobility points. The general trade shows that the increase in the network mobility results in the decrease in the EED. The EED [10] of the EC-DSR is higher at low mobility and lower at high mobility because of the selection of routes that consists with the links that has high lifetime. This decreases the link failures and results in less EED.

Packet Delivery Ratio. It is analyzed from Fig. 7, the PDR of the network decreases with the increase in the network mobility due to the increment in the mobility at the time of data transmission; link failure may occur that results in the loss of data packets that are on that active route which is broken. The PDR of EC-DSR is better when we judge against the DSR protocol due to selection of high quality route selection process in EC-DSR that results in the less re-routing processes.

Fig. 7 Average PDR versus network mobility



4 Conclusion and Future Work

Using queue load concept a routing protocol has been proposed and implemented that is an improvement on the traditional DSR protocol. It selects path with links of lowest load among the existing links so it is comparatively more successful to send data packets in congestion scenarios, medium mobility. To gauge the effectiveness in different kinds of scenarios, outcomes are created on large number of network combination with different parameters on three different network layer metrics. It is proved through simulation results that the EC-DSR is largely unaffected with increased network mobility and network load.

We can extend the proposed work by taking other parameters also into consideration that affect the link quality during the communication, such as we can take the relative mobility of two nodes from which we are selecting the link. Also, the number of hops that can be increased when searching and finding a path consists of high quality radio links. Therefore, we will try to ensure that the path length in terms of number of hops will not exceed a certain threshold. In addition to that the same technique will be tested in the future for the multicast routing protocols as they are widely used these days for many MANET applications.

References

1. Taneja, S., Kush, A.: A Survey of routing protocols in mobile ad hoc networks. *International Journal of Innovation, Management and Technology*, vol. 1, pp. 2010–0248 (2010).
2. Baghela, V., Chaudhary, A., Rajawat, S., Singh, G.: Noise Based Dynamic MANET on demand Routing Protocol for Reliable Data Transmission. In: *Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies*. pp. 1–7, ACM, Udaipur, Rajasthan, India (2014).

3. Hinds, A., Ngulube, M., Zhu, S., Aqrabi, H. A.: A review of routing protocols for mobile ad-hoc networks (manet). *International Journal of Information and Education Technology*, vol. 3(1), p. 1, (2013).
4. Bakht, H.: Survey of routing protocols for mobile ad-hoc network. *International Journal of Information and Communication Technology Research*, vol. 1(6), (2011).
5. Hamad, S., Noureddine, H., Raweshidy, H. A.: LSEA: Link Stability and Energy Aware for efficient routing in Mobile Ad Hoc Network. In: 14th International Symposium on Wireless Personal Multimedia Communications (WPMC'11), pp. 1–5, 3–7 Oct. (2011).
6. Exata Communication Simulation Plateform (EXata), <http://web.scalable-networks.com/content/exata>.
7. Tarnag, J. H., Chuang, B. W., Wu, F. J.: A Radio-Link Stability-based Routing Protocol for Mobile Ad Hoc Networks. In: *IEEE International Conference on Systems, Man and Cybernetics(SMC'06)*, vol. 5, pp. 3697–3701, 8–11 Oct. (2006).
8. Lin, X., Rasool, S.A.: Distributed Joint Channel-Assignment, Scheduling and Routing Algorithm for Multi-Channel Ad-hoc Wireless Networks. In: 26th IEEE International Conference on Computer Communications (INFOCOM 2007), pp. 1118–1126, May (2007).
9. Choudhary, S.S., Singh, V., Dadhich, R.: Performance Investigations of Routing Protocols in Manets. *Advances in Computing and Communications*, (eds) Springer, pp. 54–63 (2011).
10. Alsaqour, R., Abdelhaq, M., Saeed, R., Uddin, M., Alsukour, O., Al-Hubaishi, M., Alahdal, T.: Dynamic packet beaconing for GPSR mobile ad hoc position-based routing protocol using fuzzy logic. *Journal of Network and Computer Applications*, vol. 47, pp. 32–46 (2015).

Implementation and Integration of Cellular/GPS-Based Vehicle Tracking System with Google Maps Using a Web Portal

Kush Shah

Abstract In this paper, an efficient and inexpensive vehicle tracking system has been discussed. This device is installed in the vehicle and transmits data about its whereabouts to the server. The system contains a global positioning system (GPS) module to detect the location of the vehicle along with its speed and path. A global system for mobile communication (GSM) module and a general packet radio service (GPRS) module are also used. This is a common way to communicate with a web server; these two modules send the data of the location of the vehicle to the database where it is stored. The GPRS module ensures continuous communication with the database and helps the user track his vehicle in real-time from its website which uses the Google map API to show the position. The GSM module also allows the user to get the location in the form of short message service (SMS). This enables the user to locate the vehicle on demand without having to login to the website. This paper further describes the effectiveness and other implementations like asset tracking, of the system.

Keywords Global positioning system (GPS) • Global system for mobile communication (GSM) • General packet radio service (GPRS) • National marine electronics association (NMEA)

1 Introduction

In growing economies like India and China, the number of cars per capita is increasing dramatically the issue of security has also becomes more grave [1]. Thus, an affordable system for vehicle tracking can be useful. The vehicle tracking system is an electronic device which is installed in the vehicle to detect its GPS location. The GSM and GPS modules enable an effective and a real-time vehicle tracking to the user. The GPS module fixes the location of the vehicle and the GSM module sends the data to the web server where it is stored. A web site has been developed which

K. Shah (✉)

Charotar University of Science and Technology, Anand, Gujarat, India
e-mail: Kush289shah@gmail.com

authenticates the user and allows him to track his vehicle. The website is integrated with a Google maps API where the position of the vehicle is marked. Along with this functionality, one more service is available by SMS. The user can get the current location of the vehicle when called with the given number of the GSM module.

The Global positioning system is widely used in vehicle guidance devices which help us to navigate an effective route to a desired place. It also provides us with turn-by-turn navigation while driving. It can also provide us with the name of the streets, the speed of the vehicle, latitude, longitude, altitude, and other geographical information. It also helps us to have the knowledge of the usage of the vehicle, but this is not enough to track the vehicle remotely [2]. This same system can be used to track vehicles, but instead of using the information for navigation purposes the data could be sent to a web server where the location of the vehicle is displayed on an interactive map.

GPS is a space-based navigation and positioning system, available 24 h a day in all weather, anywhere in the world. It can be used by anyone, free of charge. GPS was conceived as a ranging system from known positions of satellites in space to unknown positions on land, sea, in air and space. GPS receivers measure time delay and decode messages from in-view satellites to determine the information necessary to complete the position and time biased calculations. This position can be expressed for example by latitude, longitude, and altitude [3].

The ubiquitous GSM provides a service to more than 500 million users throughout 168 countries worldwide [4]. This service allows the user to receive the GPS coordinates of the vehicle through SMS. The message size for this is 160 characters, which is more than enough for mentioning the latitude and longitude. Along with this GPRS module is also integrated in the system. The GPRS module is used to continuously send the information to the web server. A portal is created which receives and stores the data in a database. The GPRS module sends an HTTP request to the database and stores the information by using the GET method. This functionality is provided by the SIM908 module by Simcon. This module contains the global positioning system and the global system for mobile communication system, and by using a proper microcontroller, in this case Arduino Duemilanove, we can program the system to get the GPS coordinates and save them to the database. The system must be integrated in the vehicle itself. The antennas for the GPS and the GSM module must be integrated such that they are easily detected by the satellites.

In later parts of the paper, the study of existing systems in Sect. 2, development of the system in Sect. 3, performance and efficiency in Sect. 4 and the applications and conclusion in Sect. 5 is discussed.

2 Existing Systems

In this system, technologies like GSM and GPS are used. There are many other technologies that can be used instead of the other. Like CDMA (code division multiple access), can be used as an alternative to GSM. There are pros and cons of using both the technologies. GSM being more widely used in most of the countries

is preferable, but CDMA is also still being used extensively. GSM is used in European countries very prominently, but CDMA is still being used by some mobile carriers in the United States of America, Australia, and some Asian countries. In this system, GSM is used because it provides more flexibility over the mobile carriers. The other reason for its use is the GPRS service that it provides. It is a packet-oriented mobile data service for 2G and 3G cellular communication for GSM. In the 2G systems, GPRS provides data rates of 56–114 kbit/s. It also supports the TCP/IP protocol, which will be used to transfer the data to the web server. GPRS usage is typically charged based on the volume of data transferred and not according to the time of connection like in circuit switching, this may be in megabytes or kilobytes. This helps in keeping the cost of the system low.

The other technology used is the GPS which is developed and maintained by the Department of Defense of USA. An alternative to this is the GLONASS, which is the Russian counterpart of GPS. Both the technologies provide accurate results; the GPS is older technology and more matured. It has been around for a long time and perfected. There are more satellites supporting this technology than GLONASS. More devices support this technology and large numbers of components are available. GPS developed by the US has 31 satellites covering the planet and has been utilized in many commercial devices like the mobile phones and navigators. GLONASS has a network of 24 satellites and it is also rapidly increasing its commercial grip.

3 Development of the System

In this section, the development of the system has been discussed. The block diagram of the system is shown in Fig. 1. The main components of the system are Arduino duemilanove, SIM908 by simcom, GPS and GSM antennas, and a website with MySQLi database to store the information. SIM908 contains GPS and GSM modules both. This helps in reducing the cost and size of the device. It is connected with the Arduino serially with pins 1 and 2 which are RX and TX pins. The programs uploaded in the Arduino send commands to obtain the NMEA string through the GPS module. It contains information like latitude, longitude, number of satellites in view, time, horizontal dilution of precision (HDOP), the checksum of the string, etc. It is then manipulated using string functions to get the useful data and then sent to the database using the HTTP GET function, and other specific commands provided by the ‘SIM908 AT commands manual.’ The complete list of hardware and software used to make the system is given below.

- A. SIM908 by Simcom.
- B. Arduino Duemilanove microcontroller.
- C. GSM and GPS antennas.
- D. Arduino IDE.
- E. Wamp server.
- F. Sim card.

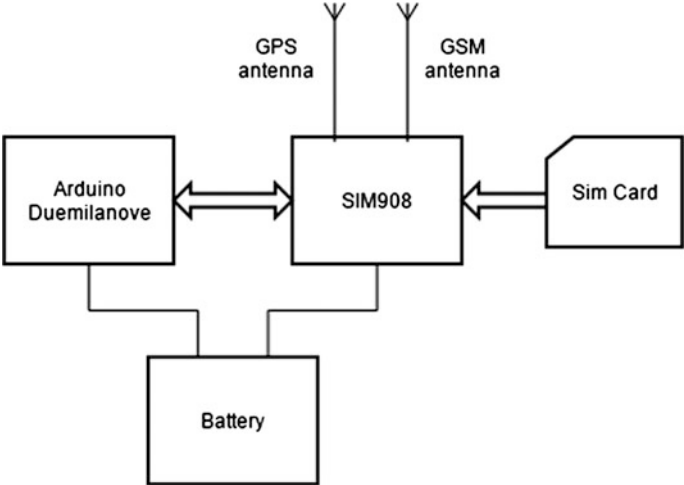


Fig. 1 Block diagram of vehicle tracking system

SIM908 module shown in Fig. 2, is a complete Quad-Band GSM/GPRS module which combines GPS technology for satellite navigation. The compact design with integrated GPRS and GPS in an SMT package significantly saves both time and cost for customers to develop GPS enabled applications. Featuring an industry-standard interface and GPS function, it allows variable assets to be tracked seamlessly at any location with signal coverage. The additional benefit of using this is the availability of resources and the fact that the entire module is very cheap. This enables us to make cheap tracking devices. The current system uses a proprietary software and hardware which makes it very expensive and this hinders the vehicle

Fig. 2 SIM908

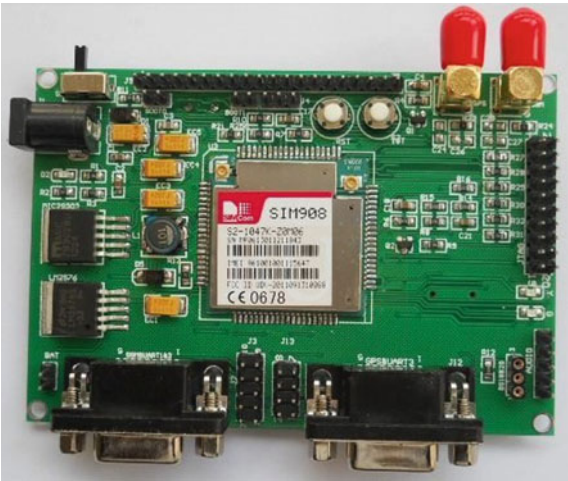
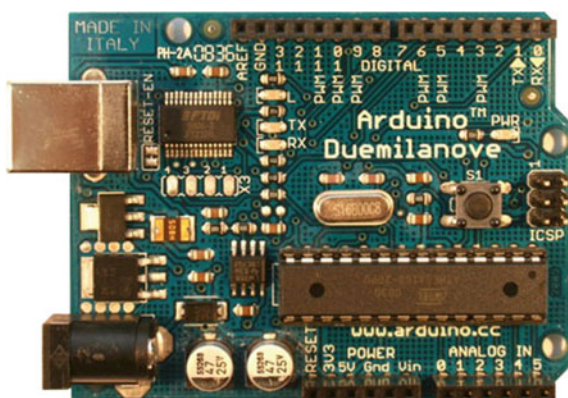


Fig. 3 Arduino Duemilanove

manufacturers to provide it along with the vehicle. The open-source modules allow it to be modified according to the users' demand.

The Arduino Duemilanove shown in the Fig. 3, is a microcontroller board based on the ATmega328, it has 14 digital inputs/outputs pins, 6 analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. After dumping the program, it simply has to be connected through a USB cable or powered it with an AC-to-DC adapter or a battery to get started.

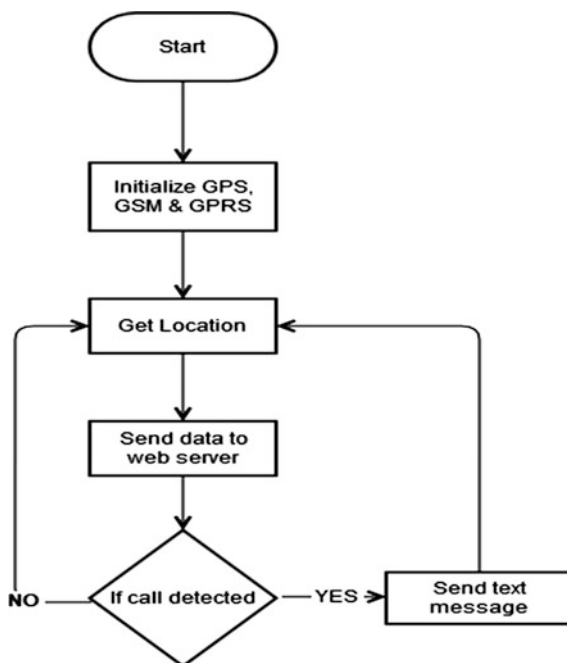
To write the program for any Arduino boards an integrated development environment (IDE) is required. It supports C and C++ programming languages. The IDE provides a user-friendly environment with a code editor with features, such as syntax highlighting, brace matching, and automatic indentation. It is also capable of compiling and uploading the program with a single click.

The Wamp server is an open-source web development platform which is used to test and run dynamic websites before uploading it to the production server. This is required to make a PHP website along with a MySQLi database to store the data from the module.

4 Performance and Efficiency

Before integrating the system with the vehicle, a SIM card must be inserted in the Sim 908 module. The mobile carrier must be chosen carefully because this system requires strong signal and GPRS to operate. The Arduino interacts with the GSM module and on powering the system the first step is to get the mobile carrier signal (Fig. 4).

The module when started initializes the GSM and GPS by sending the "AT" command and waiting for "OK." Then, it waits for the signal from the mobile carrier. The signal must be strong enough for GPRS connection to be established. Once the module has been initialized, the system waits for the GPS antenna to get

Fig. 4 Flow of the system

the signal from the satellites and determines the position. This may take a few minutes after powering up the system. Sometimes the GPS module has to be reset a couple of times to get the data. The data received by the module are in the format standardized by the NMEA (national marine electronics association). This data string contains information like latitude, longitude, altitude, speed, number of satellites being tracked, and the universal coordinated time (UTC). By using the proper string function, we can get the relevant information for our system like the latitude and longitude coordinates.

After the coordinates have been received, the APN (access point name), the username, and password of the used mobile carrier have to be entered. This sets up the GPRS connection. These coordinates must be converted to the format of decimal degrees (DDD.DDDDD°) before being sent to the web server. GET method is used to send the data and then it is stored in the database. Once, the coordinates have been formatted, the HTTP function is initialized. Then, the HTTP parameters have to be set. This requires the URL of the website. The URL along with the latitude and longitude will be set. After the URL is set, the method of transferring data must be defined, in this case GET. Therefore, AT+HTTPACTION=0 must be passed. Once the data is sent to the website, It is stored in the database and displayed on Google maps (Fig. 5).

The website is scripted in PHP and the database is based on MySQL. JavaScript and AJAX is also used to retrieve data from the server asynchronously without interfering with the display and behavior of the entire website.

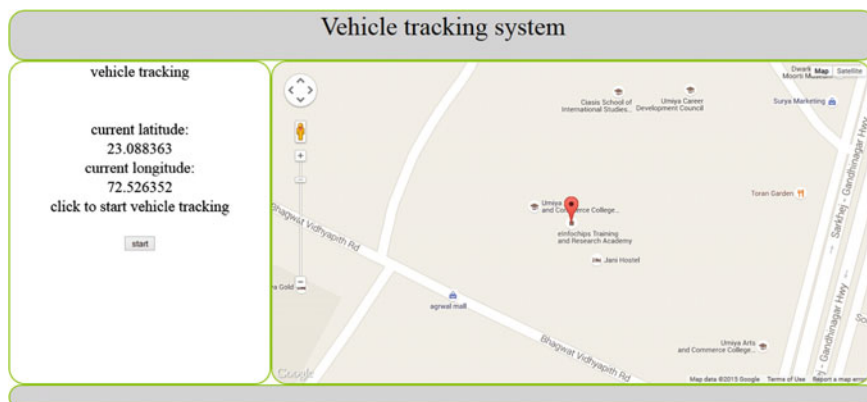


Fig. 5 Website displaying the current location

Along with these functionalities the system also waits for the user to make a phone call, and on detecting a call the system sends a text message on a predefined phone number along with the GPS coordinates. This functionality enables the user to track his vehicle without any internet connection. The mobile SIM must be maintained and kept activated for the use of this system and the user must also make sure that all services are activated.

The accuracy of the system depends on the GPS modules and the antennas which are at par with other systems. But there are a few limitations and prerequisites of this system. The GPS antenna must have LOS (Line-of-Sight) signals from the satellites moving around the earth. The module will not work properly if it is underground or in parking buildings. The mobile carrier must provide a reliable service and it must be ubiquitous. Initially, the system requires some time to detect the coordinates. This system was tested in the city area where the GSM signals are quite strong and it worked well. The system may also stop working when it enters dead zones where the signal strength is very weak or none at all.

5 Conclusion

The above discussed system is economical and applicable for daily use; it also tackles the issues of vehicle theft. As this system is an Arduino based, it is easily customizable and different functionalities can be added according to the user's requirements. Further, this system can be used for fleet management, which is necessary for transport industries. Also, it can be used by the government in public transportation and this system can be further developed into a mobile or web-based application which can be used by the people to locate their desired transport vehicle. This system only shows the real-time location of the vehicle, but it can be further enhanced to show more information like fuel status, total distance traveled

by the vehicle, number of stops taken, fuel economy of the vehicle, alternative routes, and more fuel economical routes. This system has the same functionalities as most commercialized GPS trackers minus the proprietary software and the added cost; it also provides a web portal which is platform independent, which can be accessed without any specific mobile apps. The data on the server can further be used to check the previous routes of the vehicle; this would enable the user to not only know the current position but also the previous position of the vehicle. Thus, this system has many practical uses and has the potential to be commercialized as a stand-alone product or used by car manufacturers as a device integrated into the vehicles.

References

1. http://en.wikipedia.org/wiki/List_of_countries_by_vehicles_per_capita.
2. Pham Hoang Oat, Micheal Driberg and Nguyen Chi Cuong, "Development of Vehicle Tracking System using GPS and GSM Modem", 2013 IEEE Conference on Open Systems (ICOS).
3. Xiao-Wen Chang, Introduction to GPS.
4. F Hillebrand, GSM and UMTS: The creation of global mobile communication.

Image Segmentation Using Two-Dimensional Renyi Entropy

Baljit Singh Khehra, Arjan Singh, Amar Partap Singh Pharwaha and Parmeet Kaur

Abstract Segmentation of an image is used to separate the image into several significant parts based on properties of discontinuity and similarity. Segmentation of an image is generally done with the help of thresholding technique. Thresholding is used to turn an image from gray scale to binary. The selection of suitable threshold value in the image is a challenging task. Thresholding value depends upon the randomness of intensity distribution of the image. Entropy is a parameter that is used to measure the randomness of intensity distribution of the image. In this work, Shannon-entropy-based and Non-Shannon (Renyi, Collision and Min) entropy-based approaches are used to select suitable threshold value. After this, thresholding values obtained from different approaches are tested on 6 standard test images. For evaluating, peak signal-to-noise ratio (PSNR) and uniformity (U) parameters are used. From the results, it is observed that Renyi-entropy-based approach is a better approach than other approaches.

Keywords Segmentation · Thresholding · Entropy · PSNR · Uniformity

B.S. Khehra · Parmeet Kaur
BBSB Engineering College, Fatehgarh Sahib, Punjab, India
e-mail: baljit.singh@bbsbec.ac.in

Parmeet Kaur
e-mail: pparmeetkaur@gmail.com

Arjan Singh (✉)
Punjabi University, Patiala, Punjab, India
e-mail: arjanpu@gmail.com

A.P.S. Pharwaha
SLIET, Longowal, Sangrur, Punjab, India
e-mail: amarpartapsingh@yahoo.com

1 Introduction

The method of grouping pixels of an image into homogeneous regions with regard to certain features including color, intensity, etc. is known as image segmentation. Segmentation of an image is accomplished by allocating area containing gray level under certain threshold to the background, and allocating those area containing gray levels above certain threshold to the objects, or vice versa. Thresholding converts a gray scale image into a binary image to decrease the complexity of reorganization process and reduces storage space.

$$f(x, y) = \begin{cases} 0 & \text{if } I(i, j) < T \\ 255 & \text{if } I(i, j) \geq T \end{cases} \quad (1)$$

Global threshold selection methods and local threshold selection methods are the two different classes of thresholding selection methods. In case of global thresholding method, the whole image is segmented with a single threshold value achieved with the help of gray level histogram of the image. On the other hand, local thresholding techniques divide the image into more than one sub-image and for each sub-image threshold value is decided separately. Global thresholding techniques are, in fact, very easy to execute and are computationally less complex as compare to local thresholding techniques [1].

The concept of information theoretic methodology using entropy was first introduced by Shannon [2]. The underlying principle of entropy is to utilize uncertainty for measure by which the information included in a source can be defined [3]. In information theory, there are different types of Non-Shannon entropy, such as Renyi entropy, Havrda and Charvat, Vajda, Kapur, and Tsallis entropy [4, 5]. Among the existing segmentation techniques, most of the techniques make the use of Shannon measure of Entropy [6]. The present study had evaluated two-dimensional Renyi-entropy obtained from two-dimensional histogram. Renyi Entropy is extended to get better accuracy at the same time maintaining the overall functionality. With the help of priori adjustment, texture information can be added in better manner and this will provide more accurate threshold [7, 8]. It has been conceived that Non-Shannon entropies have a better vibrant choice as compared to Shannon entropy [9, 10].

2 Entropy

A future event which cannot be observed signalizes uncertainty about the source of information. A random event, P_A which happens with probability P_A is assumed to have

$$I_A = \log[1/P_A] = -\log[P_A] \quad (2)$$

units of information. I_A is known as intellect-information of event, A , under consideration. The bottom notion of entropy in data theory is to estimate the

randomness associated with the signal or random event under consideration [11]. In case of a message which is collected of pieces and these pieces are statistically independent, then the sum total of information content of each piece will provide the information content of whole message. The collision entropy can be compiled in terms of so-called piety of a given probability distribution. It provides secondary information about the importance of specific events. Shannon entropy is used in multi-modality medical image co-registration. Renyi was able to extend Shannon entropy and it is having properties similar to the Shannon entropy [12].

$$S(t) = - \sum_{i=0}^k p_i \log_e(p_i) \quad (3)$$

Renyi Entropy is defined as:

$$R(t) = \frac{1}{1-\alpha} \log_e \left(\sum_{i=0}^k p_i^\alpha \right) \quad (4)$$

Collision entropy sometimes called Renyi entropy at $\alpha \rightarrow 2$.

$$C(t) = - \log \sum_{i=1}^L p_i^2 \quad (5)$$

when $\alpha \rightarrow \infty$ Renyi entropy converges to Min entropy. Min entropy is the smallest measure in the family of Renyi entropies and is strongest in the sense of cryptography wherein rather than making assertions about the nature of the random variable; the goal is to determine the probability that the prediction of the random variable is difficult. This main parameter in cryptography is studied using the min entropy.

$$M(t) = \min(-\log p_i) \quad (6)$$

3 Renyi Entropy to find the Optimal Threshold Value

Step1: Read input image I .

Step2: Find the size of the image I which is denoted as $[M, N]$.

Step3: Find the Histogram of input image I .

$$h_i = n_i \quad (7)$$

where n_i is the number of pixels in the image having gray level value i th, $i = 0, \dots, L$.

Step4: Find the Normalized histogram of input image I .

$$p_i = \frac{h_i}{M \times N} \quad (8)$$

Step5: Find the average of input image I .

$$k = \frac{\sum_{i=1}^M \sum_{j=1}^N I(i,j)}{M \times N} \quad (9)$$

Step6: Let $p_1, p_2, \dots, p_{k-1}$ be the probability distribution of the gray level image having value less then the average and there sum is denoted as:

$$P_{k-1} = \sum_{i=1}^{k-1} p_i \quad (10)$$

Step7: Let $p_k, p_{k+1}, \dots, p_{L-1}$ be the probability distribution of the gray level image having value greater then the average and there sum is denoted as:

$$P_t = \sum_{j=1}^t p_j \quad (11)$$

where t is the threshold value and L is 255.

Step8: Probability distribution of Foreground Region:

$$\frac{p_k}{P_t - P_{k-1}}, \frac{p_{k+1}}{P_t - P_{k-1}}, \dots, \frac{p_t}{P_t - P_{k-1}} \quad (12)$$

Step9: Probability distribution of Background Region:

$$\frac{p_{t+1}}{1 - P_t}, \frac{p_{t+2}}{1 - P_t}, \dots, \frac{p_L}{1 - P_t} \quad (13)$$

Step10: **Renyi Entropy using Foreground Region:**

$$R_F(t) = \frac{1}{1 - \alpha} \log_e \left[\sum_{i=k}^t \left(\frac{p_i}{P_t - P_{k-1}} \right)^\alpha \right] \quad (14)$$

Step11: **Renyi Entropy using Background Region:**

$$R_B(t) = \frac{1}{1 - \alpha} \log_e \left[\sum_{i=t+1}^L \left(\frac{p_i}{1 - P_t} \right)^\alpha \right] \quad (15)$$

where $\alpha \neq 1, \alpha > 0$

Step12: For Optimal Threshold value

$$T = \max_i [R_F(t) + R_B(t)] \quad (16)$$

where $i = k, \dots, L$

Step13: Segment the image using T

$$R(i, j) = \begin{cases} 0 & \text{if } I(i, j) < T \\ 255 & \text{otherwise} \end{cases} \quad (17)$$

4 Performance Measures

For evaluating the performance of Renyi-entropy-based approach, two objective measures are used. These measures are PSNR and Uniformity (U) [13]. PSNR is used to measure the equivalence between original and the binarized images. Higher PSNR shows the segmented image is better. Usually, it is easily described using the concept of mean squared error (MSE). MSE is defined as

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i, j) - R(i, j)]^2 \quad (18)$$

The PSNR is defined as:

$$\text{PSNR} = 20 \log_{10} \frac{\text{Max}_I}{\sqrt{(\text{MSE})}} \quad (19)$$

where, Max_I is the maximum probable pixel value of the given image. It is to be noted that when pixels are expressed by using 8 bits per sample, Max_I is 255 [13].

Uniformity is utilized as a measure to illustrate region similarity in a specified image. Uniformity for a threshold value T is given as follows:

$$U = 1 - \frac{2 * c * \sum_{k=0}^c \sum_{(i,j) \in R_k} (f_{i,j} - u_k)^2}{M \times N \times (I_{\max} - I_{\min})^2} \quad (20)$$

where, c represents the number of thresholds; $R_{k\text{th}}$ is k th segmented region; $f_{i,j}$ is the gray level value at (i, j) ; $M \times N$ is total number of pixels; $I_{\max}, I_{\min}$ are maximum and minimum gray level values of input image I and u_k is defined as

$$u_k = \frac{\sum_{(i,j) \in R_k} f_{i,j}}{n_k}$$

where n_k is the total number of pixels in segmented region R_k .

5 Results and Discussion

Experiments are performed on six standard gray scale test images (Usc-sipi image database) of different size and resolution. In this work, four entropy-based thresholding algorithms (Renyi-entropy based, Shannon-entropy-based, Collision-entropy-based and Min-entropy based) are implemented using MATLAB 7.7.

The original images are labeled as Figs. 1a, 2a, 3a, 4a, 5a, 6a. The histograms of these images have been prepared and shown in Figs. 1b, 2b, 3b, 4b, 5b, 6b. Figures 1c, 2c, 3c, 4c, 5c, 6c depict the Image Segmentation using Shannon based entropy; Figs. 1d, 2d, 3d, 4d, 5d, 6d represent the image segmentation using Collision-based entropy; Figs. 1e, 2e, 3e, 4e, 5e, 6e represent the image segmentation using Min-based entropy and Figs. 1f, 2f, 3f, 4f, 5f, 6f represent the image segmentation using Renyi-based entropy. These figures show that the performance of Renyi-based entropy is quite satisfactory.

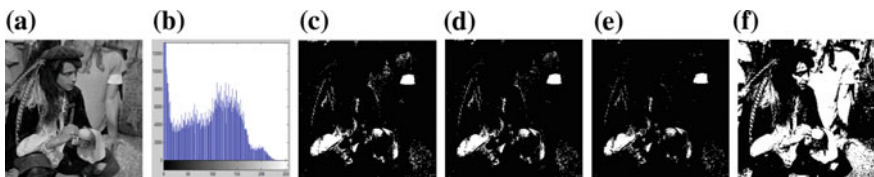


Fig. 1 a Original image, b histogram, c segmented image using Shannon entropy, d segmented image using collision entropy, e segmented image using min entropy, f segmented image using Renyi entropy

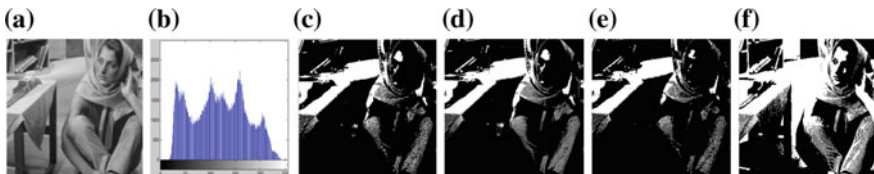


Fig. 2 a Original image, b histogram, c segmented image using Shannon entropy, d segmented image using collision entropy, e segmented image using min entropy, f segmented image using Renyi entropy



Fig. 3 **a** Original image, **b** histogram, **c** segmented image using Shannon entropy, **d** segmented image using collision entropy, **e** segmented image using min entropy, **f** segmented image using Renyi entropy

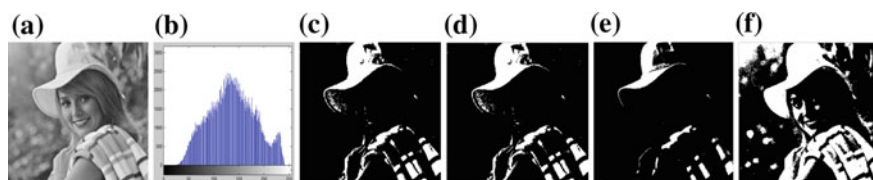


Fig. 4 **a** Original image, **b** histogram, **c** segmented image using Shannon entropy, **d** segmented image using collision entropy, **e** segmented image using min entropy, **f** segmented image using Renyi entropy

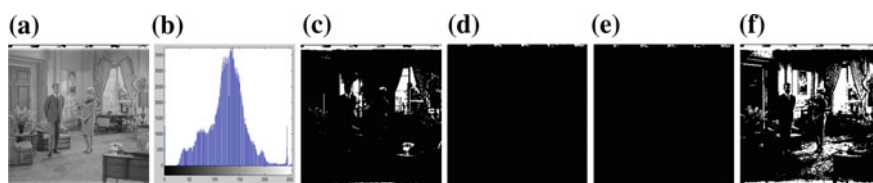


Fig. 5 **a** Original image, **b** histogram, **c** segmented image using Shannon entropy, **d** segmented image using collision entropy, **e** segmented image using min entropy, **f** segmented image using Renyi entropy

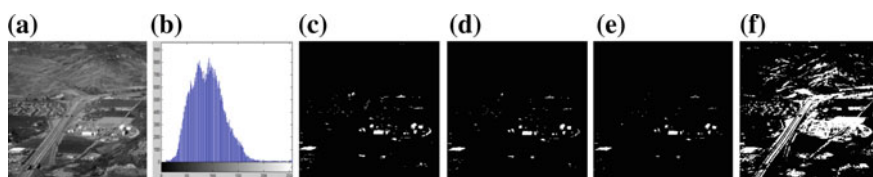


Fig. 6 **a** Original image, **b** histogram, **c** segmented image using Shannon entropy, **d** segmented image using collision entropy, **e** segmented image using min entropy, **f** segmented image using Renyi entropy

The main advantage of non-Shannon measures of entropy over Shannon entropy is that non-Shannon measures of entropy have parameters (α in case of Renyi) that can be used as an adjustable values. These parameters can play a vital role as tuning

Table 1 Parameters to evaluate the entropy using peak signal-to-noise ratio (PSNR) and uniformity (U)

S. No.	Test image	Shannon-entropy-based approach		Collision-entropy-based approach		Min-entropy-based approach		Renyi-entropy-based approach	
		PSNR	U	PSNR	U	PSNR	U	PSNR	U
1	Man1 (1024 × 1024)	8.454	0.796	8.401	0.794	8.290	0.789	9.614	0.815
2	Man (512 × 512)	7.868	0.671	7.908	0.674	7.573	0.650	9.389	0.760
3	House (512 × 512)	7.005	0.532	7.005	0.532	7.392	0.571	7.718	0.602
4	Elaine (512 × 512)	6.271	0.523	6.271	0.523	5.748	0.466	7.944	0.673
5	Couple (512 × 512)	6.718	0.608	5.920	0.536	5.919	0.536	7.518	0.673
6	Chemical plant (256 × 256)	8.648	0.735	8.601	0.733	8.562	0.731	9.077	0.749

Table 2 The optimal threshold (T) value of Shannon, collision, min, and Renyi-based entropy

S. no	Test image	T_S	T_C	T_M	T_R ($\alpha = 0.7$)
1	Man1 (1024×1024)	174	173	182	136
2	Man (512×512)	172	174	180	115
3	House (512×512)	190	190	184	178
4	Elaine (512×512)	193	193	211	157
5	Couple (512×512)	172	247	249	148
6	Chemical plant (256×256)	172	181	193	120

parameters in the image processing used for the similar set of images [4]. The PSNR and the Uniformity measure for six standard images by applying Shannon, Collision, Min, and Renyi-based entropies are shown in Table 1. The optimal threshold value of Shannon, Collision, Min, and Renyi-based entropies for the six standard images are tabulated in Table 2. From these results it is clearly evident that Renyi-based entropy has outperformed other three techniques for all the images.

6 Conclusion

In this paper, Shannon- and Non-Shannon-entropy-based image thresholding approaches are analyzed. Non-Shannon entropy-based image thresholding approaches are Renyi-entropy, Collision-entropy, and min-entropy-based approaches. The performance of these approached is tested on six standard test images. PSNR and uniformity parameters are used to evaluate the performance of these approaches. From experimental results, it is observed that the performance of Renyi-entropy-based approach is more satisfactory that other techniques for these six standard test images. From performance point of view, Renyi-entropy-based approach can be extended to colored domain and also to medical images. In general, this approach may open new directions in image segmentation research leading to interesting results.

References

1. Machta, J.: Entropy, information, and computation. In: Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts, vol. 67, no 12, pp. 1074–1077 (1999).
2. Kapur, J.: Measures of information and their Applications. In: John Wiley and Sons Publishers, New Delhi, 1st Edition, pp. 1–20 (1994).
3. Mahmoudi, L., Zaart, A.E.: A Survey of Entropy Image Thresholding Techniques. In: 2nd International Conference on Advances in Computational Tools for Engineering Applications, pp. 204–209. Beirut (2012).

4. Al-amri, S., Kalyankar, N.V., Khamitkar, S.D.: Image Segmentation by using Threshold Techniques. In: *Journal of Computing*, vol. 2, no. 5, pp. 83–86 (2010).
5. Pandey, V., Gupta, V.: MRI Image Segmentation Using Shannon and Non Shannon Entropy Measures. In: *International Journal of Application or Innovation in Engineering & Management*, vol. 3, no. 7, pp. 41–46 (2014).
6. Khehra, B.S., Singh, A.P.: Digital Mammogram Segmentation using Non-Shannon Measures of Entropy. In: *Proceedings of the World Congress on Engineering*, vol. 2. London, UK (2011).
7. Sahoo, P.K., Arora, G.: Image Thresholding using Two-Dimensional Tsallis-Havrda-Charvat Entropy. In: *Journal of Pattern Recognition Letters*, vol. 27, No. 27, pp. 520–528 (2006).
8. Chang, C.I., Du, Y., Wang, J., Guo, S.M., Thouin, P.D.: Survey and comparative analysis of entropy and relative entropy thresholding techniques. In: *IEE Proc.-Vis. Image Signal Process.*, vol. 153, no. 6, pp. 837–850 (2006).
9. Singh, A.P., Khehra, B.S.: Edge Detection in Grey Level Images based on the Shannon Entropy. In: *Journal of Computer Science*, vol. 4, no. 3, pp. 186–191 (2008).
10. Singh, A.P., Khehra, B.S.: Shannon and Non-Shannon measures of entropy for statistical Texture Feature Extraction in Digitized Mammograms. In: *Proc. of World Congress on Engineering and Computer Science*, vol. 2, pp. 1286–1291. San Francisco, USA (2009).
11. Pal, N.R., Pal, S.K.: Entropy: A New definition and its Applications. In: *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 5, pp. 1260–1270 (1991).
12. Sahoo, P.K., Arora, G.: A Thresholding method based on Two-Dimensional Renyi Entropy. In: *Pattern Recognition*, vol. 37, pp. 1149–1161 (2004).
13. Farshid, P., Abdullah, S., Sahran, S.: Peak Signal to noise ratio based on threshold method for Image Segmentation. In: *Journal of Theoretical and applied information technology*, vol. 57, no. 2, pp. 158–168 (2013).

Supervised Learning Paradigm Based on Least Square Support Vector Machine for Contingency Ranking in a Large Power System

Bhanu Pratap Soni, Akash Saxena and Vikas Gupta

Abstract In modern emerging power system many contingencies and critical operating conditions present a potential threat to system's stability. An intelligent designer at energy management center requires a paradigm which can not only predict such cases but also suggests an effective strategy for preventive control. This paper presents a least square support vector machine (LS-SVM)-based classifier to identify and rank the critical contingencies in a standard IEEE-39 bus Network (New England). This paradigm works in two stages. In first stage, the identification of two indices, i.e., voltage reactive performance index PI_{VQ} and MVA line loading index PI_{MVA} is carried out and in next stage the classification of contingencies is carried out. The proposed approach shows promising results when compared with recent contemporary techniques.

Keywords Artificial neural networks • Contingency analysis • Performance index (PI) • Static security assessment • Least square support vector machine (LS-SVM)

1 Introduction

How reliable our grid is? This question touches almost every aspect of the power system operation and control. With this thought many aspects of control, reliability, quality control, and stability are emerged in a designer's mind. As modern power

B.P. Soni (✉) · V. Gupta

Department of Electrical Engineering, Malaviya National Institute of Technology,
Jaipur 302017, India
e-mail: er.bpsoni2011@gmail.com

V. Gupta

e-mail: vgupta.ee@mnit.ac.in

A. Saxena

Department of Electrical Engineering, Swami Keshvanand Institute of Technology,
Jaipur 302017, India
e-mail: aakash.saxena@hotmail.com

system is a complex interconnected network with multiple augmentations of utilities at generation, transmission, and distribution end. Intelligent design is required at every end to ensure the secure and reliable operation of the power system. Due to competitive business environment the transmission and distribution networks are running on their stability limits. Any operating condition can present a threat to the system's stability. To predict such emergency conditions contingency ranking and screening methods are used [1–18]. Contingency ranking is a pioneer study done with the offline database of different operating conditions. This ranking is based on the calculations of the standard indices based on voltage and MVA loadings of the lines. In past various approaches are applied to predict and classify the contingencies. Artificial neural networks (ANNs) [1–3, 12], hybrid decision tree model [4], cascaded neural network [5, 7], hybrid neural network model [6], and radial basis function neural network (RBFNN) [8–10] have been applied to estimate critical contingencies and rank them in the order of occurrence and severity. Most of the approaches employed supervised learning approach for detection and understanding the complex behavior of power system under different operating conditions. Performance index (PI) has been considered as an output and responsible denominator for explaining the power system state.

Supervised learning approaches namely feed forward neural network (FFNN) [1, 2], RBFNN [8–10], cascaded neural network (CNN) [5, 7] have been presented to estimate and classify the critical contingencies for many models of power networks. The most important part of these learning approaches is input feature selection and choice of the parameters which determine the micro- and macrostructures of neural nets. In literature bus injections, state variables associated with generating and loading conditions were employed to generate a large database. To aggregate research in a more promising way, two major thrust areas are identified and are as follows: first is the development of an intelligent feature selection algorithm which can map dependent and independent variables and second is to employ fast and accurate supervised learning model to contemporary power system for accurate contingency ranking. In recent years, LS-SVM is used as a classifier in many approaches [11–13]. Huseyin et al. [11] presented a study based on wavelet transform to classify power quality events into fault events, self-regulating fault, line energizing events, and non-fault interruption events. Nine different features are extracted for this study. Similar work is reported by Sami Ekici [12] to report the power system disturbances. Power load forecasting along with ant colony optimization (ACO) is presented by Dongxiao et al. [13]. In [13] optimal feature selection is performed by ACO. Different neural topologies are presented in the work [14–18]. The size reduction of the data and optimal feature selection are the key issues addressed in these approaches.

In view of the above literature review following are the objectives of this research paper:

- (i) To develop a supervised learning-based model which can predict the performance indices for a large interconnected standard IEEE 39 bus test system under dynamic operating scenarios.

- (ii) To develop a classifier which can screen the contingencies of the power system into three states namely not critical, critical, and most critical.
- (iii) To present the comparative analysis of the reported approaches with the proposed approach based on accuracy in prediction of the PIs. The following section contains the details of the performance indices.

2 Contingency Analysis

Contingency evaluation is essential to know the emergency situations in power network. Without knowing the severity and the impact of a particular contingency, preventive action cannot be initiated by the system operator at energy management center [8]. Contingency analysis is an important tool for security assessment. On the other hand, the prediction of the critical contingencies at earlier stage, which can present a potential threat to the system stability (voltage or rotor angle) helps system operator to operate the power system in a secure state and initiate the corrective measures. In this paper line outages at every bus in New England system are considered as a potential threat to the system stability. Performance indices (PI) methods are widely used for contingency ranking [1, 6, 7, 10, 20].

2.1 Line MVA Performance Index (PI_{MVA})

On the basis of literature review it can be observed that the contingency ranking can be performed by the performance index. The system stress is measured in terms of bus voltage limit violations and transmission line over loads. System loading conditions in a modern power system are dynamic in nature and impose a great impact on the performance of the power system. An index based on line MVA flow is determined to estimate the extent of overload.

$$PI_{MVA} = \sum_{i=1}^{N_L} \left(\frac{W_{Li}}{M} \right) \left[\frac{S_i^{\text{post}}}{S_i^{\text{max}}} \right]^M \quad (1)$$

where S_i^{post} is the post-contingency MVA flow of line, S_i^{max} is the MVA rating of the line i , N_L is the number of lines in the system. In this study ($N_L = 46$), W_{Li} is the weighting factor ($=1$). $M (=2n)$ is the order of the exponent of penalty function [1]. To avoid misranking high value of exponential order ($n = 4$) is chosen in this paper. In order to classify the power system security states, on the basis of PIs calculation

0<PI<0.2	0.2<PI<0.8	0.8<PI<1
Class A (Non-Critical)	Class B (Critical)	Class C (Most-Critical)

Fig. 1 Classification criterion

the status of power system is subdivided into three categories and indicated in Fig. 1; Class A noncritical contingencies, Class B critical contingencies, and Class C most critical contingencies. Class A contingencies are not a threat to system's stability, Class B contingencies are related with the violation of the loading limits and voltage limits. However, the Class C contingencies indicate that they are not safe under any operating condition.

2.2 Line Voltage Reactive Performance Index (PI_{VQ})

The system stress is measured in terms of bus voltage limit violations and transmission line over loads. System loading conditions in an emerging power system are dynamic in nature and impose a great impact on the performance of the power system. An index based on line VQ flow is determined to estimate the extent of overload.

$$PI_{VQ} = \sum_{i=1}^{N_B} \left(\frac{W_{Vi}}{M} \right) \left[\frac{V_i - V_i^{sp}}{\Delta V_i^{Lim}} \right]^M + \sum_{i=1}^{N_G} \left(\frac{W_{Gi}}{M} \right) \left[\frac{Q_i}{Q_i^{max}} \right]^M \quad (2)$$

where $\Delta V_i^{Lim} = V_i - V_i^{max}$ for $V_i > V_i^{max}$, $V_i^{min} - V_i$ for $V_i < V_i^{min}$, $V_i S_i^{post}$ is the post-contingency Voltage at the i th bus, V_i^{max} the maximum limit of voltage at the i th bus, V_i^{min} the maximum limit of voltage at the i th bus, N_B the number of buses in the system, W_{Vi} the real nonnegative weighting factor ($=1$), $M(=2n)$ is the order of the exponent for penalty function. The first summation is a function of only the limit-violated buses chosen to quantify system deficiency due to out-of-limit bus voltages. The second summation penalizes any violations of the reactive power constraints of all the generating units, where Q_i is the reactive power produced at bus i , Q_i^{max} the maximum limit for reactive power production of a generating unit, N_G the number of generating units, W_{Gi} is the real nonnegative weighting factor ($=1$). The determination of the proper value of ' n ' is system specific. The optimum integer value ' n ' for this paper is taken as 4. In following section the basic details of least square support vector machines (LS-SVMs) are interwoven to understand the role of this supervised learning model as a regression agent and classifier.

3 Support Vector Machine

Recently, the mappings and classification problems are handled well by the artificial neural networks (ANNs). Two basic properties of neural nets make themselves different from other conventional approaches. These properties are: 1. Learning from the training samples 2. In LS-SVMs the input data is mapped with high-dimensional feature space with the help of kernel functions [11, 12]. Using kernel functions the problem can be mapped in linear form. The least square loss function is used in LS-SVM to construct the optimization problem based on equality constraints.

The least squares loss function requires only the solution of linear equation set instead of long and computationally hard quadratic programming as in the case of traditional SVMs. LS-SVM equation for function estimation can be written as shown in Eq. (3).

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad (3)$$

where α_k is the weighting factor, x are the training samples, x_k are the support vectors, b represents the bias, and N are the training samples.

The RBF kernel function for the proposed SVM tool can be written as Eq. (4).

$$K(x, x_k) = \exp\left(-\frac{\|x - x_k\|^2}{2\sigma^2}\right) \quad (4)$$

4 Proposed Methodology

Data generation is an important task in supervised learning approach. In this study a rich data of 14,000 samples is employed to train, test, and validate the networks. Following are the steps involved in the process.

- (i) A large no. of load patterns are generated by randomly perturbing the real and reactive loads on all the buses and real and reactive generation at the generator buses.
- (ii) The features are selected as per [1]. Totally 11 features as indicated in work are chosen for training purpose. These features are P_{g10} , Q_{g1} , Q_{g2} , Q_{g3} , Q_{g4} , Q_{g5} , Q_{g7} , Q_{g8} , Q_{g9} , Q_{g10} , and Q_{d14} . A contingency set for all credible contingencies are employed. $N - 1$ contingencies are the most common event in power system. Single line outages are considered for each load pattern and the value of index is stored for each iteration of the simulation.
- (iii) The obtained values of the index are normalized between 0.1 and 0.9 to train the SVM. Further the binary classification is done to train the classifier.

The system operating state contingency type and the regression performance of the network are stored for each operating scenarios. Figure 1 shows the classification criterion for the contingencies.

5 Simulated Result

The Simulink implementation of proposed approach has been carried out in MATLAB environment and tested over IEEE-39 bus test system (New England) [19, 21]. The modeling of the system and simulation studies are performed over Intel® core™, i7, 2.9 GHz 4.00 GB RAM processor unit. Bus no. 39 has been taken as slack bus. For line contingency 14,000 patterns are generated, which includes the 46 line outages and different loading patterns (300). Out of these 200 patterns are those where Newton–Raphson (NR) method failed to converge. The comparative results for the performance of the neural networks for determination of PIs are shown in Figs. 2 and 3 based on value of mean square error (MSE) and percentage R-square, respectively.

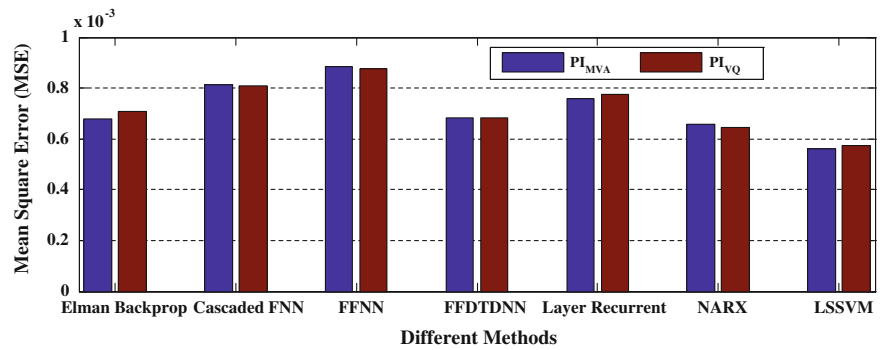


Fig. 2 Comparative performance of different neural networks (MSE)

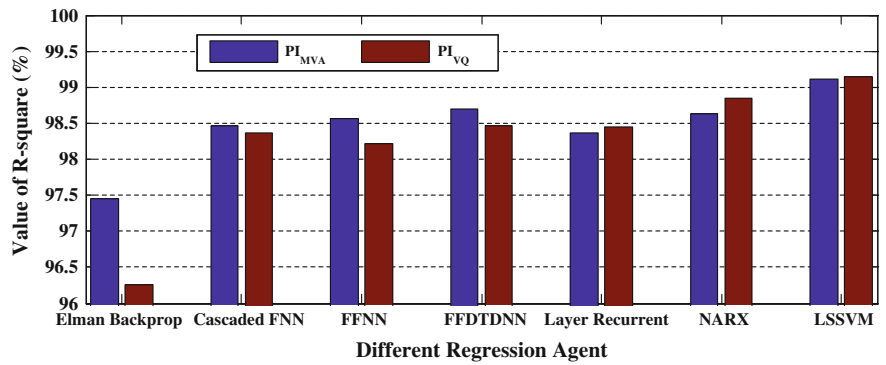


Fig. 3 Comparison of different regression agents (R-square)

Table 1 Sample result of PI_{MVA} and PI_{VQ} calculation and contingency analysis

Outage no.		1345	7811	9014	587	2984
Line no.		6–7	8–9	9–39	25–26	20–34
PI_{MVA}	NR	0.8683	0.5438	0.4971	0.4094	0.1102
	Elman backprop [14]	0.7483	0.7250	0.4635	0.3744	0.1565
	Cascaded FBNN [15]	0.7884	0.5030	0.3654	0.4649	0.1557
	FFNN [1]	0.8236	0.4457	0.3972	0.3107	0.1270
	FFDTD [16]	0.7785	0.5447	0.4482	0.3174	0.1567
	Layer recurrent [17]	0.8330	0.7349	0.3381	0.2192	0.1150
	NARX [18]	0.8177	0.3869	0.2877	0.3399	0.1465
	LS-SVM	0.8588	0.5432	0.4865	0.4014	0.1098
PI_{VQ}	NR	0.8421	0.5846	0.3876	0.4232	0.1201
	Elman backprop [14]	0.8143	0.6510	0.4231	0.3647	0.1345
	Cascaded FBNN [15]	0.8001	0.4322	0.3870	0.4515	0.1141
	FFNN [1]	0.8436	0.5561	0.4015	0.3484	0.1220
	FFDTD [16]	0.8015	0.5334	0.4312	0.3486	0.1546
	Layer Recurrent [17]	0.8451	0.5457	0.3342	0.2247	0.1340
	NARX [18]	0.8245	0.3475	0.3015	0.3846	0.1426
	LS-SVM	0.8425	0.5901	0.3870	0.4231	0.1210
Class	LS-SVM	C	B	B	B	A
	NR	C	B	B	B	A

The values of calculated indices for different contingencies as mentioned are shown in Table 1. Number of samples exhibited show the efficacy of the different methods. From Table 1 it can be judged that the line outage 6–7 during loading condition 1345 is the critical one as the values of the indices are higher for every method.

For sample no. 2984 the values of PI_{MVA} by NR method is 0.1102 and the values predicted by Elman backdrop, NARX, and cascaded FBNN are around 0.15. Higher values can clustered near the classifier boundaries and a crisp classifier will not be able to classify the state of the power system by these values.

On the other hand, the values calculated by the LS-SVM method possess lower values. It is worth to mention here that often the performance of the ranking methods is questioned due to wrong detection or misranking of a critical contingency.

The LS-SVM outperformed over the recent available topologies of neural networks (NNs) in prediction of performance indices. Classifications of contingencies are compared with the NR method and it is observed that LS-SVM can classify the contingencies well. For the ease of simplicity and understanding the excel plots are also included with the analysis. It can be observed from Fig. 2 that values of MSE for FFNN are the highest. This shows the incapability of FFNN to predict the contingencies. MSE is the residual mean square, in statistical interpolation the value closer to zero indicates that the fit is more useful for prediction. From these values it

can be concluded that LS-SVM is proven as a best regression agent for the prediction of both indices. LS-SVM method is suitable for prediction of contingencies. Values of R-square are found minimum for Elman backdrop as shown in Fig. 3. In statistical studies these values are the indication of how successful the fit is in explaining the variation of the data. The values which approach near to 1 as in case of LS-SVM shows that the machine learning model is able to predict the data very well. The value of adjusted R-square is highest in the case of LS-SVM for the calculation of both indices.

6 Conclusions

This paper proposes a supervised learning model based on least square loss function with RBF kernel function to estimate the contingency ranking in a standard IEEE-39 bus system. Following are the main highlights of this work:

- a. Comparative analysis of existing learning-based approaches for contingency ranking through standard performance indices is carried out on a large interconnected power system while considering dynamic operating conditions. It is observed that neural nets of different topologies exhibit their quality to act as a regression agent. However, the best regression results are based on MSE and R are exhibited by LS-SVM. The numerical results obtained for the indices calculation advocated the efficacy of the proposed approach.
- b. In second part the classification of the contingencies are carried out by LS-SVM. A binary classifier is obtained with three binary classes based on the values of performance indices. The performance of the SVM as a classifier is exhibited through the comparison of the results with NR method. It is concluded that SVM shows a satisfactory response to classify the contingencies.
- c. The proposed approach is suitable for online application. The operator at energy management center can easily get the details of the contingency and severity of the same with the help of these offline tested results. The study on larger system with multiple contingencies lays in the future scope.

References

1. Verma, Kusum, and K. R. Niazi. "Supervised learning approach to online contingency screening and ranking in power systems." *International Journal of Electrical Power & Energy Systems* Vol. 38, no. 1, pp. 97–104 (2012).
2. Swarup KS, Sudhakar G. Neural network approach to contingency screening and ranking in power systems. *Neurocomputing* Vol. 70, pp. 105–118 (2006).
3. Niazi KR, Arora CM, Surana SL. Power system security evaluation using ANN: feature selection using divergence. *Electr Power Syst Res* Vol. 69, pp. 161–167 (2004).

4. Patidar N, Sharma J. A hybrid decision tree model for fast voltage screening and ranking. *Int J Emerg Electr Power Syst* Vol. 8, No. 4, Art. 7 (2007).
5. Singh SN, Srivastava L, Sharma J. Fast voltage contingency screening and ranking using cascade neural network. *Electr Power Syst Res* Vol. 53, pp. 197–205 (2000).
6. Srivastava L, Singh SN, Sharma J. A hybrid neural network model for fast voltage contingency screening and ranking. *Electr Power Energy Syst* Vol. 22, pp. 35–42 (2000).
7. Singh R, Srivastava L. Line flow contingency selection and ranking using cascade neural network. *Neurocomputing*, Vol. 70, pp. 2645–2650 (2007).
8. Devraj D, Yegnanarayana B, Ramar K. Radial basis function networks for fast contingency ranking. *Electr Power Energy Syst* Vol. 24, pp. 387–395 (2002).
9. Refaee JA, Mohandes M, Maghrabi H. Radial basis function networks for contingency analysis of bulk power systems. *IEEE Trans Power Syst*, Vol 18, No. 4, pp. 772–778 (1999).
10. Jain T, Srivastava L, Singh SN. Fast voltage contingency screening using radial basis function neural network. *IEEE Trans Power Syst* Vol. 18, No. 4, pp. 1359–1366 (2003).
11. Hüseyin Erişti, Özal Yıldırım, Belkis Erişti, Yakup Demir, Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines, *International Journal of Electrical Power & Energy Systems*, Volume 49, Pages 95–103, (2013) ISSN 0142-0615, <http://dx.doi.org/10.1016/j.ijepes.2012.12.018>.
12. Sami Ekici, Support Vector Machines for classification and locating faults on transmission lines, *Applied Soft Computing*, Volume 12, Issue 6, Pages 1650–1658 (2012), ISSN 1568-4946, <http://dx.doi.org/10.1016/j.asoc.2012.02.011>.
13. Dongxiao Niu, Yongli Wang, Desheng Dash Wu, Power load forecasting using support vector machine and ant colony optimization, *Expert Systems with Applications*, Volume 37, Issue 3, Pages 2531–2539 (2010).
14. Toha, Siti Fauziah, and M. Osman Tokhi, MLP and Elman recurrent neural network modelling for the TRMS, In: 7th IEEE International Conference on Cybernetic Intelligent Systems, pp. 1–6. (2008).
15. Goyal, Sumit, and Gyandera Kumar Goyal, Cascade and feed forward back propagation artificial neural networks models for prediction of sensory quality of instant coffee flavoured sterilized drink, *Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition* Vol. 2, Issue 6, pp. 78–82. (2011).
16. Ghosh, Atish K., and David L. Lubkeman, The classification of power system disturbance waveforms using a neural network approach, *IEEE Transactions on Power Delivery*, Vol. 10, Issue 1, pp. 109–115. (1995).
17. Williams, Ronald J., and David Zipser, A learning algorithm for continually running fully recurrent neural networks, *Neural computation* Vol. 1, Issue 2, pp. 270–280. (1989).
18. Lin, Tsungnam, Bil G. Horne, Peter Tiño, and C. Lee Giles, Learning long-term dependencies in NARX recurrent neural networks, *IEEE Transactions on Neural Networks*, Vol. 7, Issue 6, pp. 1329–1338. (1996).
19. Power system test cases at <http://www.ee.washington.edu/pstca/>.
20. J.C.S. Souza, M.B. Do Coutto Filho, M.Th. Schilling, Fast contingency selection through a pattern analysis approach, *Electric Power Systems Research*, Volume 62, Issue 1, Pages 13–19 (2002), ISSN 0378-7796, [http://dx.doi.org/10.1016/S0378-7796\(02\)00016-0](http://dx.doi.org/10.1016/S0378-7796(02)00016-0).
21. Matpower 4.0 User's Manual: <http://www.pserc.cornell.edu/matpower/manual>.

Process Flow for Information Visualization in Biological Data

Sreeja Ashok and M.V. Judy

Abstract Every day new discoveries are made in the field of molecular biology and genetics and the sheer volume of data coming out of scientific journals are overwhelming. To collect process and integrate this raw and complex information is probably the most challenging task of current generation of academicians and research scholars. Creating a combined platform by integrating various forms of biological data like DNA sequences, protein structures, or metabolic pathways helps bioinformaticians and computational biologists for efficient data analysis. Current work proposes a structured process flow by integrating different data exploration techniques and visualization techniques that aid in visual extraction of information from biological data.

Keywords Dimension reduction • Visualization • Clustering • Classification • Gene expression • Biological networks

1 Introduction

Visualization is an approach that tones down terabytes of random information freely available into small packages that can be read, processed, and reproduced in different ways. The net result is to allow the users to coherently explore large quantity of information in a short period of time. There are various visualization tools available in computational science today that has benefited academicians and corporate worldwide. Its increasing significance in multiple domains can be attributed to the fact that although large quantities of data are available to the public, only little of it can be actually viewed at a time. When presented as text, it is very

Sreeja Ashok (✉) • M.V. Judy
Department of Computer Science & I.T., Amrita School of Arts & Sciences,
Amrita Vishwa Vidyapeetham, Amrita University, Kochi, India
e-mail: sreeja.ashok@gmail.com

M.V. Judy
e-mail: judy.nair@gmail.com

difficult to examine the millions and billions of data items. Data visualization can be an answer in such instances, where any information, be it numerical or alphanumeric can be represented in a map; chart, bars, pie chart, etc., to compare and study. Biological data are heterogeneous and highly complex in nature. It involves both clinical and molecular data. A systematic approach is necessary for discovering underlying structure, detecting trends and patterns, and deriving true actionable insights from these huge dataset. Here we provide a summary of visual data exploration techniques and propose a work flow for data exploration in biological data that integrates both conventional mining methods and information visualization techniques.

1.1 Visual Analysis of Gene Expression Data

Gene expression is the series of steps from transcription to translation that controls the amount of gene products (proteins or RNA) synthesized. It occurs in all organisms, extremely regulated within the body and is crucial for the functioning of the cell. The mechanism is altered in different cells and tissues to meet their specific metabolic demands and carry out processes specific to the organs. Because of such tight scrutiny, gene expression is precisely regulated using various transcriptional and translational machineries [1]. Recent advances in technology have enabled large-scale measurement of gene expression; usually it is done with the objective of comparing different cell types, like normal cells and those in disease state. The advantage here lies in the fact that data for thousands of genes can be obtained under varying conditions with manual analysis. Commonly used techniques for measuring gene expression include northern blotting, qPCR, and expensive DNA microarrays. In DNA microarray technology, the cDNA derived from the mRNA of known genes is attached to the plates. The sample contains genes from both the normal as well as the diseased tissues. Spots with high intensity are obtained for diseased tissue gene if the gene is overexpressed in the diseased condition. This expression pattern of both is then compared.

1.2 Visual Analysis of Biological Networks

Biological networks constitute interplay of pathways and processes of all bimolecular data like DNA, RNA, and proteins and their interactions. The relational patterns of the characteristics and the relations in the dataset can be found out using multi-relational data mining techniques. Biological networks are better represented using a network graph consists of vertices and edges. Different types of biological networks are included in Table 1.

Table 1 Biological networks

Biological network	Description	Tools
Metabolic networks	Network of biochemical reactions which involves enzymes, substrates, and the products synthesized during each step of the pathway and regulation of the relative amounts of metabolites involved	KEGG, BiNA
Regulatory networks	Gene expression modulation, transcriptional regulation	BiNA, cytoscape
Protein–protein interactions networks	Protein functionality depends on the nature of its interactions with other proteins, metabolites, or molecules	iPfam
Signaling networks	System of communication among the molecules for normal functioning and maintenance of cellular integrity	SBML

2 Visualization Process Flow

Data mining and visualization techniques work hand in hand to enable complete elucidation and user interpretation of large datasets. The sandwich technology that combines both techniques disciplining their respective limitations are particularly useful in the context of molecular biology where large volumes of sequences and gene arrays can be efficiently represented in graphs, trees, and chains. Visualization process of molecular data focuses on data preprocessing, data reduction, clustering/classification, analysis and knowledge discovery. The process flow is shown in Fig. 1.

Data Preprocessing Data that contribute noise will lead to wrong analysis and results. Preprocessing techniques help in converting raw data to meaningful biological data by removing noise, low intensity, bad quality, and empty spots from biological data using normalization, filtration, sampling, extraction, labeling, scanning, etc.

Dimensionality Reduction Methods It is an effective approach to downsize the data. For example, when the dataset has thousands of genes and few samples and the objective is to classify novel samples into known disease type, dimensionality reduction methods help in finding a subset of informative genes which can be processed for further analysis. Different methods include

- **Principal Component Analysis**—Transforms the attribute set to a minimal set of principal components that explain the main variations of the data. It is used to visualize high-dimensional profiles as projections in lower dimensional spaces (usually two dimensional, sometimes also three dimensional). There is always a loss of information in the process; goal is to minimize the loss of information. PC view, graphs, and scatter plots are few data representation charts used. PC view is a line graph that represents the sum total of principal components, eigenvectors, and expression values. The first 2 or 3 principal components play

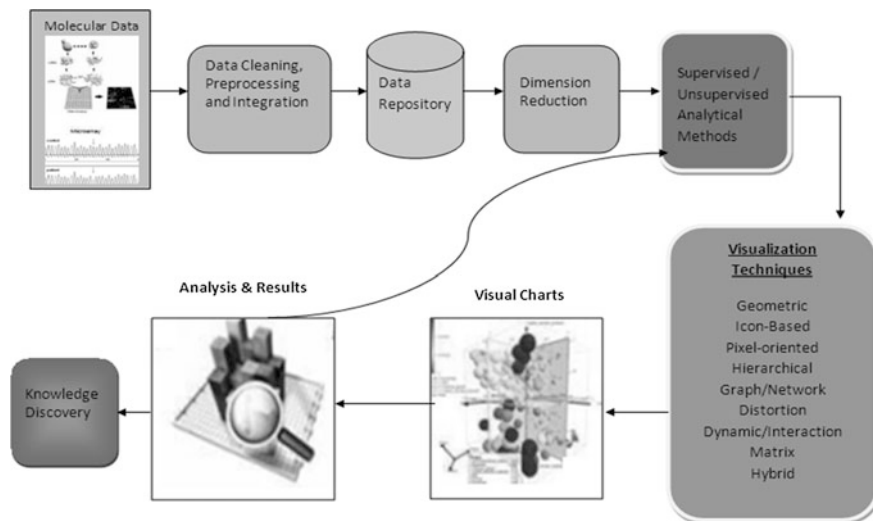


Fig. 1 Visualization process flow

significant role in feature reduction. Colors can be changed to represent the core attributes. Graphs and scatter plots can be used to represent the vital number of PCs [2, 3].

- **Parallel Coordinates**—Includes both clustering and visualization techniques where biclusters can be detected and visualized. There is a possibility of missing data when there is a large and complex pattern of data. Overplotting is another major challenge. Parallel coordinate plots are used for data representation [4].
- **Partial Least Squares**—Supervised method for constructing core components by maximizing the covariance between the dependent variable Y and the predictor variables Xs . Benefit of PLS dimension reduction is the opportunity to visualize the data by graphical representation [5].
- **Sliced Inverse Regression**—Supervised approach where the response information is utilized in achieving dimension reduction. Efficient approach for both attribute reduction and visualization in high-dimensional dataset [6].
- **Factor Analysis**—Determines a set of unnoticeable common factors that explain the major discrepancies of the data. The initial attribute dimensions are linear combinations of the factors derived. Factor model plots are used to represent the model [7].
- **Multidimensional Scaling**—The coordinate axes in multidimensional space represents the similarity or dissimilarity of the data. Different distance measures like Euclidean, Manhattan, Minkowski, etc., are used to compute the similarity of the data items [8].
- **Fastmap**—Fast and efficient algorithm for clustering and visualization. It is an iterative process to reduce the attribute size while preserving the distances as much as possible. Helps in analyzing medical datasets where one-dimensional

data, e.g., ECGS, two-dimensional data, e.g., X-rays and three-dimensional images like MRI brain scans are used for pattern mining and predictions [9].

Classification Helps in understanding the complex relationship/interaction among the various conditions and features of a biological object. For example, a training dataset has diseased and normal cells and when a new cell is obtained, classification process has to automatically determine whether it is normal or a diseased cell. Classification techniques [10–12] that are generally used in biological data analysis are detailed in Table 2.

Table 2 Classification techniques

Technique	Pros	Cons	Visualization techniques
Discriminant Analysis. Classified as Linear discriminant analysis (LDA), Diagonal quadratic discriminant analysis (DQDA), Diagonal linear discriminant analysis (DLDA) based on nature of class densities	Multiple dependent variables, reduced error rates, easier clarification of between-group differences	Patterns of correlations between variables are considered to be equivalent from one group to the next, the relationships between variables are taken to be linear in all groups, multicollinearity, extremely sensitive to outliers	Scatter plot Data summary
Single decision tree includes C4.5, CART, decision stump, random tree, REPTree ensemble decision tree includes bagging, AdaBoost, ADTree, random forests	Generate understandable knowledge structures, low computational cost, can handle symbolic and numeric input variables, can identify the important attributes	Instability, difficulties in branching the trees when number of samples is too low	Tree-like graph
Artificial intelligence approaches—probabilistic induction (Naive Bayes method)	Simplicity, computational efficiency, good performance. performs well when the number of predictors (Xs) are very large	Requires large no. of records to obtain good results. In the absence of predictor category in training data, Naive Bayes assumes zero probability for that category of the predictor biased results	Nomogram—graphical representation of numerical relationships
Artificial intelligence approaches—artificial neural network (ANN)	Efficient problem solving, massive parallel processing, reprogramming is not needed since ANN learns by itself	Needs training to operate, has to emulate architecture of microprocessors, high processing time for large networks, do not have a built-in variable selection mechanism hence there is need for careful consideration of predictors	Network maps, parallel coordinates plot (PCP)—for multidimensional visualization

(continued)

Table 2 (continued)

Technique	Pros	Cons	Visualization techniques
Similarity based methods—nearest neighbor analysis	Simplicity and lack of parametric assumptions, perform surprisingly well In the presence of a large training set, especially when each class is characterized by multiple combinations of predictor values	Execution time to obtain the nearest neighbors in a large training set can be too expensive, curse of dimensionality	Heat map, scatter plot, histogram
Max-margin classifiers —support vector machine	Produce very accurate classifiers. Less over fitting, robust to noise	Computationally expensive, thus runs slow	Univariate histogram, scatter plot, line chart, bidimensional graph with PDCOLOR coding

Clustering Based on the homogeneity of data, datasets are grouped using different clustering algorithms. Microarray has been a standard approach for representing biological data. Each column in the microarray gene expression represents a condition and each row represents a gene. Clustering microarray data helps in making hypothesis about potential functions of genes, protein–protein interactions, etc. Different clustering techniques that are commonly used for biological data analysis [13–15] are represented in Table 3.

Table 3 Clustering techniques

Technique	Pros	Cons	Visualization techniques
Hierarchical clustering	Number of clusters not required in advance, input parameters—choice of the (dis) similarity, computes a complete hierarchy of clusters, intuitive algorithm, good interpretability	May not scale well: runtime $O(n^2 \log n^2)$, no precise clusters: a “flat” partition can be derived afterwards, no automatic discovering of “optimal clusters”, susceptible to outliers, when tree is big, interpretation is difficult	Dendrogram, Treemap view, gene tree, array tree, matrix tree plot or two-way dendrograms
k-means—partitioning clustering for numerical large dataset	Efficient: $O(xyz)$, where z is no. of objects, y is the no. of # clusters, and x is the no. of iterations, easy implementation,	Cluster size needs in advance, responsive to outliers, clusters formed are convex shaped, cluster results are dependent on the	Spherical shape, scatter plot, heat map—elegant graphical representations of cluster contents

(continued)

Table 3 (continued)

Technique	Pros	Cons	Visualization techniques
	simplified gaussian mixture model, normally get nice clusters	initial partition, local optimum solutions	
PAM—partitioning clustering for numerical large dataset	Easy and simple to understand and execute, fast convergence within little iteration, less sensitive to outliers, using basic dissimilarity functions of objects	Different initial sets of medoids can lead to different clustering results, multiple iterations suggested with different initial sets of medoids, clustering results depends on the units of measurement, standardization of variables is necessary for varying variable magnitude and nature	Arbitrary shape
DBSCAN—density based clustering for numerical high-dimensional dataset	Better performance for low dimensional data, input parameters required are MinPts and Eps	Sampling would affect the density measures, not partitionable for multiprocessing systems	Arbitrary shape
SOM—model based clustering method- numerical low dimensional dataset	Different distance measures and joining criteria to form big cluster, clusters has interpretation on 2D geometry	Very heuristic algorithm. Suboptimal solution due to 2D geometry restriction	Network graph, table-node weight and edge weight, linked brushing
Fuzzy c-means	A data point can be in multiple clusters, more natural representation of the behavior of genes, genes usually are involved in multiple functions	User defined membership cutoff, no “natural” visualization of the data, “outlier” genes forced to belong to some cluster	Convex hulls, scatter plot, 3-D plots, histogram
CLICK—tight clustering	Allow genes not being clustered; only produce tight clusters, ease the problem of accurate estimation of # of clusters, Biologically more meaningful	Slower computation when data large	Heat map

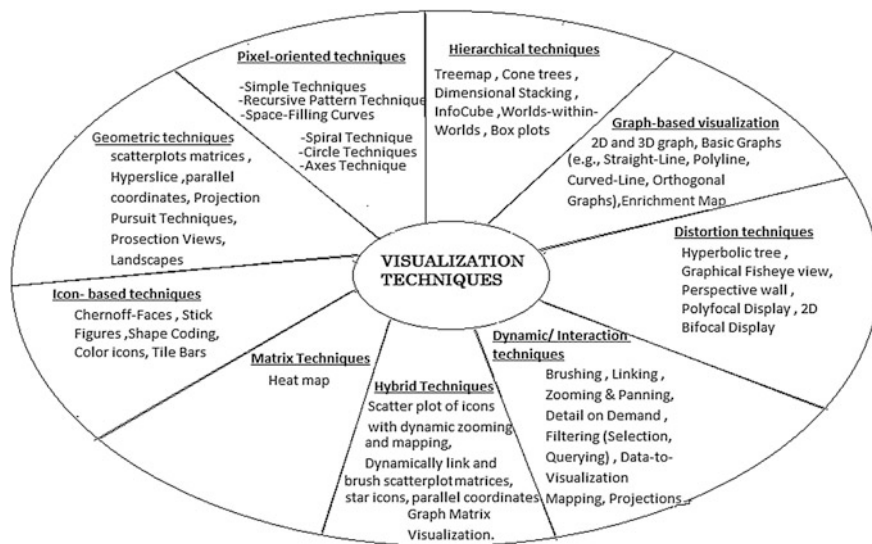


Fig. 2 Visualization techniques

3 Summary of Visualization Techniques

Visualization techniques help in visually inspecting and interacting with two-three-dimensional view of processed dataset. Figure 2 shows the summary of visualization techniques commonly used [1, 16].

4 Conclusion

Research journals spew out vast quantities of heterogeneous, dynamic, and largely unprocessed information that has to be transformed into a coherent and user-friendly format easily accessible to all. Exploring and analyzing these huge volume of biological data has become more and more difficult. Visualization is important because of the increased scope and complexity of the nature of biological studies where new specialized fields are continually emerging. Analytical methods together with visualization techniques play a major role in exploring, analyzing, and presenting meaningful inferences.

Acknowledgments This work is supported by the DST Funded Project, (SR/CSI/81/2011) under Cognitive Science Research Initiative in the Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham University, Kochi.

References

1. Tangirala Venkateswara Prasad and Syed Ismail Ahson, (2006): "A survey of Visualization of microarray gene expression data", *Bioinformation*, pp. 141–145.
2. Hotelling, H. (1933). "Analysis of a Complex of Statistical Variables with Principal Components." *Journal of Educational Psychology*, 24: 498–520.
3. Abdi, H., & Williams, L.J. (2010). "Principal component analysis." *Wiley Interdisciplinary Reviews: Computational Statistics*, 2: 433–459.
4. A. Inselberg., (2009), *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
5. Boulesteix, A.-L. and K.Strimmer (2006). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics* 7, 32–44.
6. Li, L.: (2010) Dimension reduction for high-dimensional data. In: *Statistical Methods in Molecular Biology*, vol. 620, pp. 417–434. Springer Protocols.
7. de Tayrac, M., Le, S., Aubry, M., Mosser, J., & Husson, F. (2009). Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. *BMC Genomics*. 10, 32.
8. Strickert M, Teichmann S, Sreenivasulu N, Seiffert U (2005). High-Throughput Multi-Dimensional Scaling (HiT-MDS) for cDNA-Array Expression Data. *Lecture Notes in Computer Science*, 3696, 625–634.
9. C. Faloutsos and K.I. Lin. FastMap, (1995): A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *ACM SIGMOD International Conference on Management of Data*, pages 163–174.
10. C.J.C. Burges, (1998) "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167.
11. A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and Z. Yakhini, (2000): "Tissue Classification with Gene Expression Profiles," *J. Computational Biology*, vol. 7, pp. 559–584.
12. N. Friedman, M. Linial, I. Nachman, and D. Pe'er, (2000): "Using Bayesian Networks to Analyze Expression Data," *J. Computational Biology*, vol. 7, pp. 601–620.
13. Banfield, J. and Raftery, A. (1993). "Model-based Gaussian and non-Gaussian clustering". *Biometrics*, 49, 803–821.
14. Ben-Dor, A. and Yakhini, Z. (1999): "Clustering Gene Expression Patterns", *Proceedings of the 3rd Annual International Conference on Computational Molecular Biology (RECOMB 99)*, pp. 11–14, Lyon, France.
15. Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta, (2012): "A Comparative Study of Various Clustering Algorithms in Data Mining," *International Journal of Engineering Research and Applications (IJERA)*, vol. 2, Issue 3, pp. 1379–1384.
16. Ashraf S. Hussein (2008): "Analysis and Visualization of Gene Expressions and Protein Structures", *Journal of Software*, vol. 3, no. 7.

A Performance Analysis of High-Level MapReduce Query Languages in Big Data

Namrata Singh and Sanjay Agrawal

Abstract The current era is an era of big data analytics. One of the challenges of big data is mining of the relevant data out of huge volume of databases where the data is present in variety of formats. MapReduce is providing a viable solution to analyze this type of data, but it has some limitations and weaknesses too. Hence, the high-level query languages have evolved for querying massive amount of data over MapReduce. In this research paper, the authors have analyzed the performance of the three prominent high-level query languages viz. Pig Latin, HiveQL, and JAQL based on the query processing time. We have first stored data in the Hadoop distributed file system, processed the data for wordcount, and web log processing benchmarks and then analyzed it. An experimental analysis of the three languages has been performed on unstructured data format by doubling the size of the dataset.

Keywords High-level query languages • Pig • Hive • JAQL • Hadoop • Big data

1 Introduction

The current era is an age of digital revolution. The emerging trend toward the digital services and technology is to digitize every minute information. With the growth of the internet, global communication, and networking has increased. As a result, the need of storage, transmission, and accessing this information or data has become very significant. Over the past few years, there has been tremendous increase in the volume of data. This has given rise to the term big data. Big data has been widely used to describe about the exponential growth of the data with respect to variety, volume and velocity and thus has become one of the major areas of research and analytics

Namrata Singh (✉) · Sanjay Agrawal
Department of Computer Engineering and Applications, National Institute
of Technical Teachers' Training and Research, Bhopal, India
e-mail: namratacs0027@gmail.com

Sanjay Agrawal
e-mail: sagrawal@nittrbpl.ac.in

now-a-days. The key contributors to the growth of this data are the internet, social media, sensors, smart phones, etc. This data needs to be stored and processed. The traditional storage and processing mechanisms like the relational database management systems have failed to process this large amount of data. This big data problem is now being handled by various technologies like NoSQL databases [1], Hadoop [2], etc. These technologies provide an effective platform for dealing with the enormous amount of data, which needs to be effectively gathered, processed, and analyzed.

Among them, Hadoop is one of the technologies which can be used to deal with various types of data. Since the data is originating from various domains, analytics has become a great challenge for big data. This data is very valuable and acts as a crucial component in analysis as the organizations are using this data for their growth. Querying the relevant and important data out of the dataset is a crucial task. For this, many query languages have been built on top of Hadoop so as to skip the burden of writing the MapReduce programs for processing the data in Hadoop. High-level query languages (HLQLs) have evolved as the Hadoop query languages to provide minimum customization to the programmer.

2 Related Research

HLQLs provide a means to query the pertinent data of importance from the datasets. These HLQLs have the specifications of SQL and are the dataflow languages in Hadoop. Some of these languages which are used for querying are Pig, Hive, and JAQL. These three scripting languages have different features and strengths and are used for the processing of different data formats. These have got different computational power and processing capacities. In the literature [3], the authors have performed an analysis on a database of five lakh records using Pig, Hive, and JAQL to make the access of the results in an efficient, fast, and easy way whenever a query is made to the database. They have acquired the data from social media using Flume. Data is analyzed using mapreducers in Pig, Hive, and JAQL. Some researchers [4], have done a comparative analytical study between Pig, Hive, and JAQL with the help of various parameters. These three technologies are used for intelligent decision making and massive data analytics. In the literature [5], the authors have proposed a big data platform built on Hadoop MapReduce, gluster file system, Apache Pig [6], Apache Hive [7] and JAQL [8]. A systematic performance comparison between the three HLQLs has been made in [9] using scale up, scale out, and runtime metrics. The working of the three languages has been provided in the lab manual by IBM [10].

3 MapReduce and Hadoop

Mapreduce is a software framework for writing applications which process vast amounts of data on large clusters of nodes. It is also known as a programming model and as an associated implementation for processing and generating large



Fig. 1 The figure shows the working of the mapreduce framework. The input and output are both in the form of key-value pairs

datasets. MapReduce job splits the input dataset into individual chunks and then send for parallel processing to the map and reduce tasks.

The map function takes as input the list of unstructured records and emits for each a set of intermediate key-value pairs. For each key, list of values are produced by map libraries which are applied as input to the reduce operation. Then the reduce libraries collate these values and merge into smaller set of values or a single value. Figure 1 shows the working of the mapreduce framework.

Hadoop is a framework for storage and processing of large datasets on clusters of nodes so-called “commodity hardware.” Its core components consist of Hadoop Distributed File System (HDFS) [11] and Mapreduce [12]. This framework was designed so as to automatically handle the hardware failures. HDFS provides high speed I/O access to data, fault tolerance and guarantees high reliability of the system.

4 High-Level Query Languages

HLQLs are those languages which are constructed on top of Hadoop to provide more abstract query facilities in comparison to the low level languages such as Java. A number of HLQLs have been designed out of which the most important are the three namely Pig Latin, HiveQL, and JAQL.

As the core technology of the Hadoop is the mapreduce parallel processing model, all of the HLQLs which run on Hadoop are the mapreduce-based query languages. Programs written in these languages are compiled into a sequence of mapreduce jobs. These three different technologies make it easier to write mapreduce programs in Hadoop. These high-level languages help us to write programs that are smaller than their equivalent Java code. All these languages translate high-level languages into mapreduce jobs so that the programmer can work at a higher level other than writing mapreduce jobs in Java or any other lower-level languages supported by Hadoop.

4.1 Pig

Pig is a high-level platform for writing the mapreduce programs which are used with Hadoop with a much higher level of abstraction. The language used by this

platform is Pig Latin. Pig Latin is an abstraction of the mapreduce programming framework which makes it a high-level query language on top of Hadoop. It can be extended using user-defined functions (UDFs) in which the external code can be written in java, python, javascript, ruby, etc.

It provides a data flow interface for Hadoop for data summarization and advanced querying. It is one of the components built on top of HDFS which is meant for processing of huge amount of data with the help of multiple transformations.

4.2 *Hive*

Hive is one of the components of Hadoop on top of HDFS which provides a data warehouse infrastructure for Hadoop. It provides an SQL dialect, called Hive query language (HiveQL) for querying the data stored in the Hadoop cluster. The features of Hive are SQL-like since it provides the functions like group-bys, aggregation, joins, etc.

The hive architecture [13] is mainly composed of five main components. The first one is the user interface which provides an interface for the users to submit their queries. The next one is the driver which receives the queries and does session handling and provides APIs based on JDBC/ODBC interfaces. The third one is the compiler which parses the query and performs semantic analysis and generates an execution plan. Fourth one is the metastore which performs the validation of the RDBMS schema or query. It is the internal database of Hive which maintains the metadata information of the tables. The last one is the execution engine, responsible for executing the execution plan generated by the compiler.

4.3 *JAQL*

JAQL is a data processing and query language used for JSON query processing on big data. JAQL programs run in the JAQL shell. The main goal of JAQL is the manipulation of the semi-structured data. JAQL consists of many built-in operators and functions. It also consists of many core expressions operated on nested arrays.

5 Experimental Environment

In our experimental setup, we have worked in a Hadoop single node cluster environment. Hadoop is an open source project of Apache Software Foundation and is freely downloadable from the apache website. It is installed on the Linux file system.

The hardware and software specifications are mentioned below in Table 1.

Table 1 Hardware and software specifications

Parameter	Specification
CPU	CPU W3565
Speed	3.20 GHz
RAM	12 GB
Operating system	CentOS 6.3 64-bit
Disk space	30 GB
Software	Hadoop 1.0.4
Component	Pig 0.10.0, Hive 0.12.0, JAQL 0.5.1

6 Analysis of Experimental Results

6.1 Benchmarks for Analysis

The following two benchmarks are used for analysis:

The wordcount benchmark Wordcount benchmark is one of the standard benchmarks for analysis purpose as found in the literature review. The first experiment is based on the analysis of the query processing time of the three components on a given dataset of text files which are in unstructured format.

The web log processing benchmark The web log processing is another standard benchmark used for the analysis purposes. The second experiment is based again on the query processing time of the three components on a given dataset of web log files which are in common log format.

6.2 Analysis of Query Processing Time Based on Wordcount Benchmark

The first experiment performed is based on the analysis of the query processing time based on the wordcount program as the standard benchmark. Here we have used an unstructured format dataset of text files [14]. We have downloaded the dataset and performed the analysis of the query processing times on the three components viz. Pig, Hive, JAQL. While analyzing these components both in Pig and Hive the dataset is used as it is in downloaded format but while doing analysis in JAQL it is converted in Java Script object notation (JSON) format. The property of JAQL is that it processes the data which are in JSON, CSV, XML formats, etc. Therefore, while doing analysis in JAQL, we have first converted the textual data into JSON format and then analyzed it.

For performing the analysis, the unstructured data is first uploaded in the HDFS for all the three components of Pig, Hive, and JAQL. Then, the processing of the data is done to get the required results. The wordcount program has been written for all the three components in the languages viz. Pig Latin, HiveQL, JAQL. The queries which are written in these languages retrieve the count of words in the textual data.

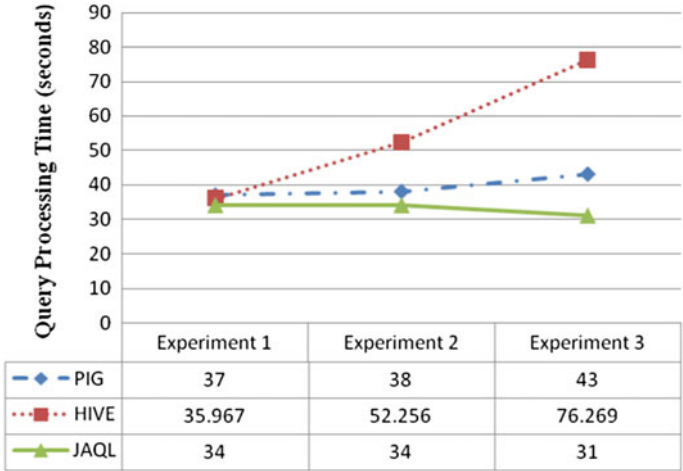


Fig. 2 Figure displaying the variation in the query processing times of Pig, Hive, and JAQL

The output query results show the CPU time and the query processing times which are analyzed. We have done the analysis by performing three experiments upon doubling the size of the files as shown in Fig. 2.

After getting the results from the above processing of data, the parameter on the basis of which we have done the analysis is the query processing time. All the three components have their own way of processing the query since they are all different languages built for different purposes. The analysis also depends upon the data format used. Pig and Hive use the unstructured text data format and JAQL uses the JSON data format. The text data has been converted into JSON format to be processed by the JAQL.

The dataset size has been doubled in every experiment and analyzed for query processing time for all the three components. The observation shows that Pig outperforms Hive in comparison among both but JAQL shows the least query processing time among all the components. Therefore, our analysis shows that in processing the textual data which is in unstructured format, JAQL is the best. Since the data is in JSON format, therefore it easily processes the data. Then, Pig processing shows that when we increase the dataset size, the query processing times gradually increase. Hive takes largest time for query processing therefore it is not much suitable for analysis of unstructured data.

6.3 Analysis of Query Processing Time Based on Web Log Processing Benchmark

The second experiment performed is based on the analysis of the query processing time based on the weblog files as the standard benchmark. The weblog dataset [15]

is downloaded from the NASA website. This data is in common log format where the details of the web log viz. IP address, HTTP code are present. We have loaded the data into the HDFS and performed the analysis of the query processing times on the three components viz. Pig, Hive, JAQL. While storing these components in HDFS both in Pig and Hive the dataset is used as it is in downloaded in common log format but in JAQL it is converted in (Java Script Object Notation) JSON format. Therefore, we have first converted the web log data into JSON format and then stored it. The whole data for all the three components of Pig, Hive and JAQL is first uploaded in the HDFS and then analyzed.

Then, the processing of data is required to get the required results. We have processed the data to find the number of times the given IP address has been accessed or the hit counts of the IP address. The program for finding the number of hits has been written for all the three components in the languages viz. Pig Latin, HiveQL, JAQL. The queries which are written in these languages retrieve the number of hits as the result.

After getting the results from the processing of the data, the parameter on the basis of which we have done the analysis is the query processing time. All the three components have their own way of processing the query since they are all different languages built for different purposes. The analysis also depends upon the data format used. Pig and Hive use the common log format which by default is the format of the web log server data format and JAQL uses the JSON data format. The web log data has been converted into JSON format to be processed by the JAQL. The dataset size has been doubled in every experiment and analyzed for query processing time for all the three components.

A comparative analysis from Fig. 3 shows that Hive and JAQL outperform PIG in comparison among the three and Hive shows the least query processing time among all the components. Therefore, our analysis shows that in processing the

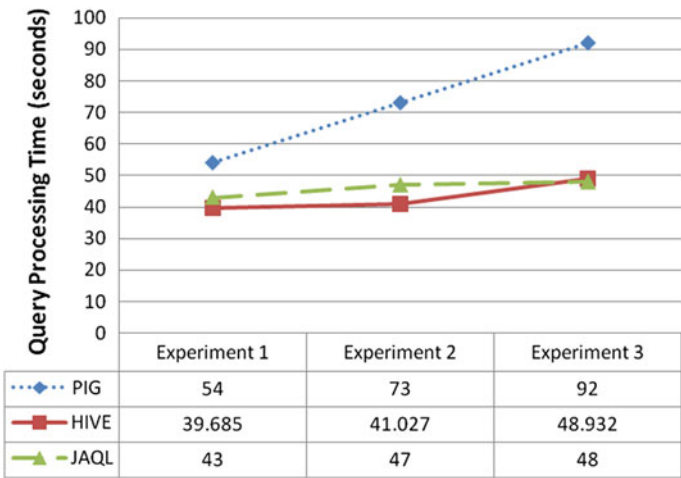


Fig. 3 Figure showing the variation in the query processing time of Pig, Hive, and JAQL

web log data which is in unstructured format, Hive is the best. Also JAQL shows performance equivalent to the Hive. Since the data is in JSON format, therefore it easily processes the data. Pig takes the largest time for query analysis therefore it is not much suitable for analysis of web log data.

7 Conclusion and Future Work

In this research paper, we have presented the performance analysis of the three high-level query languages viz. Pig, Hive, and JAQL. This analysis has been done with two benchmarks namely wordcount and web log processing. The parameter used for analysis is the query processing time. While analyzing the text data, JAQL proves to be the best since the stored data is in JSON format. Comparing between Pig and Hive, Pig outperforms Hive with second least query processing time in text data analysis. While analyzing the web log data, Hive shows the least query processing time among all the three components and Pig takes the highest time, therefore Pig is unsuitable for analysis of web log processing data. JAQL also shows similar query processing time to Hive since the data is in JSON format. In future, the authors are intended to analyze the query languages in multinode cluster environment with some more parameters and variety of datasets.

References

1. NoSQL, <https://en.wikipedia.org/wiki/NoSQL>.
2. Apache Hadoop, https://en.wikipedia.org/wiki/Apache_Hadoop.
3. Shetha, M., Mehtab, P., Deulkar, K.: Enhancing Massive Data Analytics with the Hadoop Ecosystem. *Int. J. of Engineering And Computer Science* 3, 9061–9065 (2014).
4. Rathee, S.: Big Data and Hadoop with components like Flume, Pig, Hive and Jaql. In: *International Conference on Cloud, Big Data and Trust 2013*, pp. 78–82, Bhopal (2013).
5. A Comparison of Big Data Analytics Approaches Based on Hadoop MapReduce, https://www.academia.edu/3502325/A_Comparison_of_Big_Data_Analytics_Approaches_Based_on_Hadoop_MapReduce.
6. Pig (programming tool), [https://en.wikipedia.org/wiki/Pig_\(programming_tool\)](https://en.wikipedia.org/wiki/Pig_(programming_tool)).
7. Apache Hive, https://en.wikipedia.org/wiki/Apache_Hive.
8. Jaql, <https://code.google.com/p/jaql/>.
9. Stewart, R.J., Trinder P.W., Loidl H-W.: Comparing high level mapreduce query languages. In: Temam, O., Yew, P.-C., Zan, B. (eds.) *APPT 2011, LNCS*, vol. 6965, pp. 58–72. Springer, Heidelberg (2011).
10. IBM Software Information Management HDFS, MapReduce, Pig, Hive, and Jaql Hands-On Lab, https://www.ibm.com/.../files/.../04_Hadoop_Core_Lab.pdf.
11. HDFS Architecture Guide, http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
12. MapReduce Tutorial, http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html.
13. Design, <https://wiki.apache.org/confluence/display/Hive/Design>.
14. Reuters-21578 Text Categorization Collection Datasets for single-label text categorization, <http://www.cs.umb.edu/~smimarog/textmining/datasets/20ng-train-all-terms.txt>.
15. NASA-HTTP, <http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>.

Variance-Based Clustering for Balanced Clusters in Growing Datasets

Divya Saini, Manoj Singh and Iti Sharma

Abstract k -Means is a very popular clustering algorithm. We modify its objective to achieve a clustering method which produces more balanced clusters. The proposal can be adapted in a framework where dataset keeps growing and number of clusters is decided within the algorithm to achieve balanced clustering. This is done without affecting the time complexity. Experimental results are in favor of the proposal.

Keywords Data mining · k -Means · Clustering · Balanced clusters · Dynamic datasets

1 Introduction

Clustering processes attempt to partition data into groups of similar objects [1]. This is helpful not only in data analysis but also in many other applications involving compression or grouping. That is why clustering is a very active research area. Among so many clustering techniques, k -means [2] ranks in the top ten popular algorithms [3]. It is still most researched due to inherent flexibility. Changing requirements of data analysis need incremental clustering algorithms or adaptive techniques, which can change the parameters according to growing dataset. But these efforts largely come through machine learning methods or genetic algorithms. Besides soft computing, research to make the existing simple algorithms adaptive has not been much. Moreover, the incremental approaches tend to preserve the previous results and the computations are centered on the newly added data plus the experience gained from previous data. Instead, the approach could be of considering all data as fresh dataset and reconsidering entire analysis. This ideal approach would then cost much effort and time. Thus, it does not suit the actual need. What we aim to

Divya Saini · Manoj Singh
Gurukul Institute of Engineering, Kota, India

Iti Sharma (✉)
Rajasthan Technical University, Kota, India
e-mail: itisharma.uce@gmail.com

propose is modifications to a classic clustering method such that each time new data arrive some parameters can be adjusted so that only necessary changes are done in previous clusters and new data become a part of the previous cluster structure. Specifically, we modify k -means algorithm, with ' k ', the number of clusters, as the adjusting parameter. The objective function of the heuristic is slightly changed as in [4] and serves an extra purpose of decision criteria to increase the value of ' k '.

The underlying idea of the proposal is of penalizing large clusters to ensure balance. The approach of regularizing size of clusters according to the relative winning frequency of their representatives, that is, how many points they attract, is used in [5, 6]. These methods start from random centers and in subsequent steps avoid adding new points to those clusters which are already much populated. Some other analogous strategies of producing balanced clusters are proposed in [7, 8].

Besides big data where data is voluminous and increases with high velocity [9], certain other applications need that algorithms learn value of ' k ' instead of deciding it a priori. Applications like image processing, coverage problem of sensor networks, data compression, etc., need the clusters to be balanced. This paper proposes a clustering method aimed at producing balanced clusters. The value of k is changed until the desired level of balance is reached.

Next section describes briefly the concept of balancing clusters through a modified objective function in k -means, as suggested in MinMax k -means [4]. Section 3 discusses the proposed approach. Experimental setup and results are described in Sect. 4.

2 MinMax k -Means

MinMax k -means [4] is an approach to tackle k -means initialization problem by altering its objective. The method starts from a randomly picked set of centers and tries to minimize the maximum intra-cluster variance instead of the sum of the intra-cluster variances. For proper initialization, a weight is associated with each cluster, such that clusters with larger variance are allocated higher weights, and a weighted version of the sum of the intra-cluster variances criterion is derived. The choice of cluster weights directs the effort towards minimizing those clusters that currently have large variance. Thus, the resulting cluster structure in the end does not have clusters with large variance. Moreover, the weights are learned automatically during the iterative procedure of cluster assignment. A parameter which decides the degree of penalty towards large variance clusters is an input to the algorithm.

The MinMax k -means [4] method has two alternate phases of minimization and maximization. Minimization phase is like traditional k -means which attempts to minimize the distance of a data point from a centroid. Maximization phase is for computing the weights using closed-form expressions. The purpose of the maximization step is to adjust bad effects of initialization, thus, guaranteeing good solutions even with bad seeding. Overall, the minmax k -means method produces balanced clusters in terms of variance.

3 Proposal

The idea of modifying the objective function towards having low variance clusters can give a good heuristic. Authors in [4] have used this to nullify the effects of bad seeding. We propose to use the concept of penalizing the clusters according to variance in order to achieve balanced clusters as an output in cases where dataset keeps growing. We first propose a clustering method in which ‘ k ’ is input. The purpose is to have the most balanced clusters, as indicated by maximum variance and sum of variances. Thereafter, we proceed to apply this method to dynamic data. Whenever the dataset grows, it is expected that the value of maximum variance should not deviate more than an allowed margin. This can be achieved by increasing value of ‘ k ’ until the desired variance is achieved.

The penalty imposed on cluster with largest variance should be adjusted in such a way that the further iterations move towards balancing. We propose to do this by removing that data point from such cluster which is causing high variance and assigning it to the smallest cluster. This serves twofold purpose. First, in the next iteration, points will get rearranged causing more balance. Second, for the dataset which are highly imbalanced and are in a risk of having empty cluster if not initialized properly, this approach will never produce empty clusters. The proposed method is briefly outlined as follows:

ALGORITHM 1: Basic Variance-based Clustering

INPUT: Data set $X = \{\mathbf{x}_i\}_{i=1}^N$, number of clusters k

OUTPUT: Cluster assignments $\{c_i\}_{i=1 \dots N}$ and centroids $\{\mathbf{m}_j\}_{j=1 \dots k}$

Step 1: Select initial centroids, $\{\mathbf{m}_j\}_{j=1 \dots k}$ at random

Step 2: Assign datapoints to clusters according to minimum distance $c_i = \operatorname{argmin}_{1 \leq j \leq k} \|\mathbf{x}_i - \mathbf{m}_j\|^2$

Step 3: Repeat steps 4 to 10 until convergence

Step 4: If (current_maximum_variance < previous_maximum_variance) then Step 5

Step 5: Assign data points to clusters according to minimum distance

Step 6: Else Step 6 to 9

Step 7: Select the cluster with largest variance

Step 8: Pick the data point with maximum distance from the centroid in this cluster

Step 9: Assign this point to cluster with smallest variance

Step 10: Update centroids as mean values of data points

The measurement of variance for Step 4 requires some computation substeps. The intra-cluster variance is measured through sum of the distances of the objects from the centroid within a cluster. Then, the weight attributed to a cluster is decided as the measure of the cluster's contribution towards overall variance. Mathematically put, intra-cluster variance for each cluster is computed as

$$v_j = \sum_{\forall \mathbf{x}_i \text{ in cluster } j} \|\mathbf{x}_i - \mathbf{m}_j\|^2 \quad (1)$$

where $\|\mathbf{x}_i - \mathbf{m}_j\|$ denotes the distance between point $\mathbf{x}_i$ and its centroid $\mathbf{m}_j$. If Euclidean distance is used, then $\|\mathbf{x}_i - \mathbf{m}_j\|^2 = (x_{i1} - m_{j1})^2 + (x_{i2} - m_{j2})^2 + \dots + (x_{iD} - m_{jD})^2$. D denotes the number of attributes. The weight of a cluster is computed as

$$w_j = \frac{v_j}{\sum_k v_j} \quad (2)$$

Further, we compute such values which could indicate the shape of clusters, i.e., the weighted sum of variances of a cluster as

$$\varepsilon_{w_j} = \sqrt{w_j} v_j \quad (3)$$

The proposed method attempts to minimize the sum of the weighted sum of variances, that is

$$\varepsilon_w = \sum \varepsilon_{w_j} \quad (4)$$

Hence, Step 4 of the algorithm compares previous value of ε_w with the current value of ε_w . In Step 7, the cluster with largest variance is selected as the cluster ' l ', such that $l = \operatorname{argmax} \varepsilon_{w_j}$. Step 8 picks a point $\mathbf{x}_r$, where $r = \operatorname{argmin} \|\mathbf{x}_i - \mathbf{m}_j\|^2$. The cluster to which $\mathbf{x}_r$ is reassigned is picked in Step 9 as $q = \operatorname{argmin} \varepsilon_{w_j}$.

This basic clustering method can now be used for clustering datasets that keep growing. The idea is to decide an allowable threshold, above which if there is change in value of ε_w , k needs to be increased. Thus, the algorithm can be outlined as

ALGORITHM 2: Adaptive Variance based clustering

INPUT : clustered dataset $X = \{\mathbf{x}_i\}_{i=1}^N$, cluster assignments $\{c_i\}_{i=1\dots N}$, centroids $\{\mathbf{m}_j\}_{j=1\dots k}$, new data $Y = \{\mathbf{x}_i\}_{i=1}^M$, threshold δ

OUTPUT : Cluster assignments $\{c'_i\}_{i=1\dots N+M}$, number of clusters k' and centroids $\{\mathbf{m}_j\}_{j=1\dots k'}$

Step 1 $k' = k$

Step 2 $\{c'_i\}_{i=1\dots N} = \{c_i\}_{i=1\dots N}$

Step 3 Compute ε_w over X using Eq (1) to (4), store as `previous_var`

Step 4 Append dataset X with Y , so that now the data is $X = \{\mathbf{x}_i\}_{i=1}^{N+M}$

Step 5 Assign new datapoints of X to clusters according to minimum distance $c_i = \operatorname{argmin}_{1 \leq j \leq k'} \|\mathbf{x}_i - \mathbf{m}_j\|^2$, $N+1 \leq i \leq N+M$

Step 6 Compute ε_w over entire dataset X using Eq (1) to (4), store in `current_var`

Step 7 While (`current_var - previous_var >  $\delta$`)

Step 8 Increment k' by 1

Step 9 `previous_var = current_var`

Step 10 Assign clusters of all data points using **Algorithm 1** with inputs X and k'

Step 11 Compute `current_var` as ε_w over dataset X using Eq (1) to (4)

Step 12 end while

The proposed modification has the benefit that it does not affect the linearity of runtime of the algorithm. Also, it can be used effectively when data arrives in chunks and emphasis is purely on clustering rather than classification. It can be used for other similar applications of k -means.

4 Evaluation

Experiments with popular datasets are performed to ensure that the proposed algorithm performs better than MinMax k -means [4], in terms of variance of resulting clusters. Experiments on synthetic dataset were conducted to see how the algorithm adapts itself to increase the value of ‘ k ’.

Iris dataset has been used for widely experimenting for many clustering algorithms. It has 150 instances with 4 attributes, which can be grouped into 2 or 3 clusters on a priori basis. We have conducted the experiment with increasing number of values of k from 2 till 6, and studied its effect on cluster quality as indicated by value of ε_w and maximum variance that is maximum ε_{w_j} . Table 1 shows the recorded values for the proposed algorithm and standard k -means.

The performance of the proposed method is better than standard k -means. The difference is not high for iris dataset due to the low range of values and less number of instances.

In order to compare with MinMax k -means algorithm, experiments over *Escherichia coli* and dermatology dataset were conducted. *E. coli* (UCI) [10] includes 336 objects which are proteins from the *E. coli* bacterium and have 7 attributes. It is a highly imbalanced dataset and hence a good candidate to compare two different clustering methods. Dermatology (UCI) [10] has records of 366 patients who suffer from six different types of skin disease. Each record contains 34 features, clinical and histopathological. This dataset is also unbalanced. Table 2 records the results for both. Since, the methods proposed in [4] have an extra parameter, we have used the entire range of results reported. As can be observed the proposed method gives better clusters for both datasets.

Table 1 Results for iris dataset

No of clusters	k -means		Proposed algorithm	
	$\max \varepsilon_{w_j}$	ε_w	$\max \varepsilon_{w_j}$	ε_w
2	111.59	123.95	111.6	124.03
3	28.298	48.08	26.682	47.75
4	29.729	45.36	10.689	29.24
5	11.27	25.02	8.35	23.3
6	6.97	17.609	6.005	18.5

Table 2 Comparison of proposal with MinMax k -means

Method	Output	Dataset	
		<i>E. coli</i> ($k = 4$)	Dermatology ($k = 6$)
MinMax k -means [4]	$\max \varepsilon_{w_j}$	5.29–4.80	1513.85–1368.05
	ε_w	15.94–15.72	5703.26–5672.82
Proposed Variance-based clustering	$\max \varepsilon_{w_j}$	4.376	1145
	ε_w	10.951	5210

5 Conclusion

MinMax k -means was proposed to circumvent the ill effects of bad initialization in k -means. We have used this concept to propose a variance-based clustering algorithm. The proposed algorithm itself is a better performer than standard k -means and MinMax k -means in terms of balanced clusters. It does not produce empty clusters. It can be used with random seeding, thus saving time of expensive initialization.

Also, we have shown how it can be adapted to be used with growing datasets. Whenever a dataset is appended with new records, the proposed algorithm iteratively clusters the new points within the previous clusters and increases the number of clusters if required. This property of the proposed method makes it suitable for many applications of data compression and image processing.

6 Future Scope

Further research can be done on employing the optimization techniques with the proposed algorithm as has been done previously with k -means only. Actual use of the proposed method in application areas like image processing, feature selection, area coverage in sensor networks etc., needs to be explored. Extending the proposal for categorical and mixed data is also under research. Using other metrics to measure cluster quality in place of variance can also produce interesting results.

References

1. Everitt, Brian S., Landau, Sabine, Leese, Morven, Stahl, Daniel: Cluster Analysis, 5th Edition, Wiley (2011).
2. Lloyd, S.: Least squares quantization in pcm. Information Theory, IEEE Transactions, Vol. 28 (2), pp. 129–137 (1982).
3. Wu et al, X.: Top 10 algorithms in data mining, Knowledge and Information Systems, Vol. 14 (1), pp. 1–37 (2008).
4. Tzortzis, G., Likas, A.: The MinMax k -Means clustering algorithm Pattern Recognition Vol 47, pp 2505–2516, Elsevier (2014).
5. Wang, C.-D., Lai, J.-H., Zhu, J.-Y.: A conscience on-line learning approach for kernel-based clustering, International Conference on Data Mining (ICDM), pp 531–540 (2010).
6. Banerjee, A., Ghosh, J.: Frequency-sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres, IEEE Transactions on Neural Networks 15(3), pp 702–719 (2004).
7. Bradley, P.S., Fayyad, U.M.: Refining initial points for k -means clustering, International Conference on Machine Learning (ICML), pp. 91–99 (1998).
8. Su, T., Dy, J.G.: In search of deterministic methods for initializing k -means and Gaussian mixture clustering, Intell. Data Anal. 11(4), pp 319–338 (2007).
9. Big Data Taxonomy, by Big Data Working Group (2014) Available at <https://cloudsecurityalliance.org/research/big-data>.
10. Frank, A., Asuncion, A.: UCI machine learning repository (2010) URL <http://archive.ics.uci.edu/ml>.

Conceptual Framework for Knowledge Management in Agriculture Domain

Nidhi Malik, Aditi Sharan and Jaya Srivastava

Abstract Abundance of information has its challenges also. In this digital age, there are enormous sources of information. Information is scattered and available in different formats. Users have to struggle a lot in order to find out the desired information. This information needs to be structured and delivered in a manner which fulfills the requirements of the user. Agriculture is an important domain of Indian system as well as Indian economy. However, knowledge management in the context of Indian agriculture domain is very poor. So, efforts should be made to manage knowledge so that it can be launched on the Web and should have the provision of being converted to semantic web, e-commerce site etc.

Keywords Knowledge • Knowledge management • Ontology • Semantic web

1 Introduction

Knowledge is a very general term. Technically, it refers to understanding a subject whether it is theoretical or practical. It may be acquired through experience, education, or learning [1]. There are many processes involved in acquiring and then applying knowledge to systems. Specifically to a domain, knowledge is the information about it and it is further used to solve problems. For problem solving, this information needs to be represented in some form. In order to use this information in an efficient way, it is important to manage knowledge in such a way that it highlights the concepts, perspectives, context, etc. Doing so will enable the use of semantic

Nidhi Malik (✉) • Aditi Sharan
Jawaharlal Nehru University, New Delhi, India
e-mail: nidhimalik14@gmail.com

Aditi Sharan
e-mail: aditisharan@mail.jnu.ac.in

Jaya Srivastava
Indian Institute of Technology Delhi, New Delhi, India
e-mail: jaya@iitd.ac.in

information which in turn will boost up the overall performance of related applications. In this work, we are proposing a framework for knowledge management in agricultural domain. The subdomain of fertilizers has been chosen as the case study for this work. Most of the content available is in the form of unstructured data that does not fit neatly into structured form. Considering the effort and money that goes into creating such volumes of data, there is a compelling need for development of systems which manages information for farmers, researchers, policy makers, etc.

Knowledge management is important in all fields such as enterprise, business, e-learning, government, agriculture, airlines, railway, military, etc. Knowledge management is essential for using knowledge effectively and efficiently at right time [2]. The reasons for managing knowledge are:

1. There are different sources of information. They are stored in varying formats/platforms. This disparity affects the efficient usage of knowledge.
2. Knowledge from different sources is inconsistent and redundant. This in turn leads to incorrect information being used by people involved.
3. There is lack of coordination between different units/tasks which is primarily because of inefficient ways of representation and management of knowledge.

Ontologies have become the best choice as the medium of knowledge management in recent years for a range of computer science applications including the Semantic Web, health sector, tourism, bioinformatics, etc. The basic characteristics of ontology that prove to be beneficial in knowledge management are:

1. The foundation concepts of an ontology are defined and specified with respect to the domain itself. So, it becomes easy to infer further meaningful knowledge. It is possible to define meaningful relationships among the concepts in ontology. The relationships and properties that can be assigned to concepts will make the semantic basis stronger.
2. Since Ontology acts as a repository for organizing and managing knowledge, it provides functionalities for querying, reasoning, and inferring information.
3. Ontologies enable integration of information sources, sharing, and reusing knowledge.

Rest of the paper is organized as follows. Section 2 discusses the need for knowledge management in agriculture domain, Sect. 3 describes the proposed architecture, Sect. 4 gives implementation details and Sect. 5 gives the conclusion.

2 Need of Knowledge Management for Agriculture Domain

India is a developing country and agriculture plays a very important role in its economy. It is among the top two farm producers in the world. Over 70 % of the rural households depend on agriculture. Agriculture contributes about 17 % to the total GDP and provides employment to over 60 % of the Indian population [3].

With the extensive use of technology in all fields, agricultural sector is also being modernized. There are large number of projects/research going on in different aspects and directions. None of the projects can be successful without making use of technology and knowledge. It is becoming complex to manage the amount of knowledge that is being generated nowadays [4]. Ontologies have proven to be capable of efficiently managing information because of the basic structure and properties that they have. [5] discusses that the most important role of ontology in knowledge management is to enable and enhance knowledge sharing and reusing. Moreover, it provides a common mode of communication among the agents and knowledge engineer.

With Semantic Web, the semantics of what the user is trying to find out comes into play as the software agents are made more “intelligent” while carrying out the required task given by the users rather than just going about the keywords in the user query like the case with conventional search engines. Ontology makes up for a model which plays an important role in the implementation of such type of systems. The structure of information captured by ontologies can be used to share the common understanding of domain among people and software agents. For example, there are various different websites that contain information on fertilizers or provide fertilizer recommendations. Now, if there is a common underlying ontology that these websites share, the software agents can extract and aggregate information in a more efficient way. This aggregated information can then be used to answer user queries.

With respect to agricultural domain, fertilizers management forms an important subdomain of agriculture along with (not less than) crop management, soil management and pest management. However, when we observed the state of art regarding information management, we realized that of the above-mentioned subdomains, fertilizer is most poorly represented. It is always treated as dependent entity, generally in terms of crops and soil. So, we have chosen to develop an ontology for fertilizers and also querying and reasoning on this ontology. The knowledge base has been created by referring different resources from NCAP, IARI, etc.

3 Proposed Work

The aim of this work is to show how ontologies can be used for managing domain knowledge. The work is divided into three steps. We have proposed a conceptual framework of a knowledge management system. In the second step, we have developed an ontology for fertilizers, and the third step briefly shows how this ontology can be accessed by real time users. The step by step working is discussed below.

3.1 Conceptual Framework

The figure gives the conceptual framework for the overall process. The working process revolves around two tasks: the creation of fertilizers ontology,

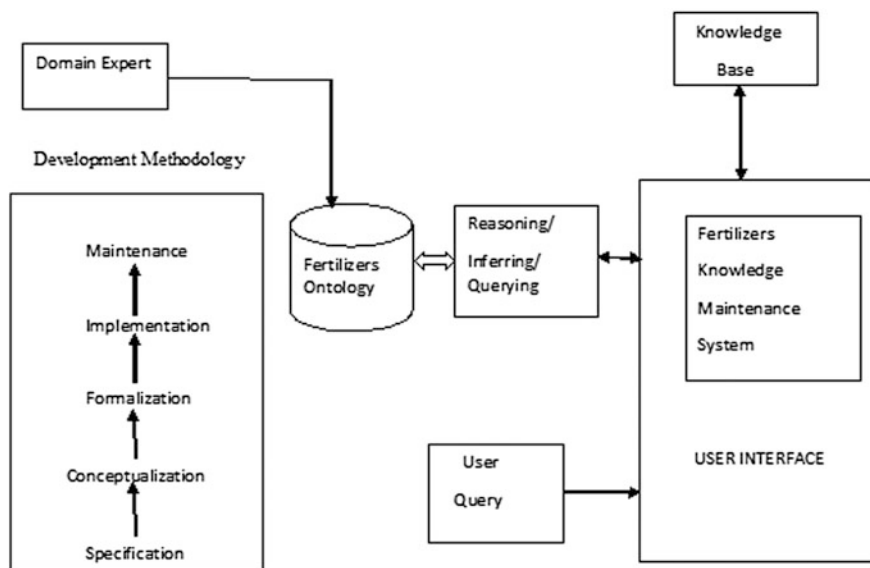


Fig. 1 Conceptual framework of the proposed system

querying/reasoning over it and making it accessible via an interface. The system has been operational and created ontology can be made available on request since it is further expanded. The methodology that we have followed to develop this ontology is a generalized phase by phase procedure of the activities that are required for ontology development (Fig. 1)

Much help is also taken from the Ontology Development 101 [6] for the small details in the ontology conceptualization phase. Ontology development is an iterative process and the ontology can be updated continuously. As of now, we have included static information in fertilizers such as type of fertilizers, nutrient contents, residual effect on the soil, preferred soil type, time of application, method of application, etc.

3.1.1 Ontology Development

A lot of information on agriculture is available on the web. However, they are scattered and unorganized and cannot be accessed efficiently. In order to have a meaningful extraction of the information from the large corpus of documents present in the web, there should be a shared understanding of the domain and developing domain ontology for the same addresses the problem. Ontology in simplest terms can be called as a collection of entities and their relationships with each other.

There are many definitions given for an ontology in the literature but the one that is most suited is given by [3]. “Ontology is formal specification of a *conceptualization*”. Ontology can play a critical role in representing knowledge for a domain. More specifically

A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

A number of stages are involved in the process of ontology development. The usually accepted stages involved in developing an ontology are specification, conceptualization, formalization, implementation, and maintenance. Reference [7] proposed the following activities to be performed during the various stages of ontology development (Fig. 2).

The view of the ontology that we have created is shown in Fig. 3. The activities shown above are all done while development of this ontology. The following steps given by [6] are followed for ontology construction:

- Step 1 Determine the domain and scope of the ontology
- Step 2 Consider reusing existing ontologies
- Step 3 Enumerate important terms in the ontology
- Step 4 Define the classes and the class hierarchy
- Step 5 Define the properties of classes—slots
- Step 6 Define the facets of the slots
- Step 7 Create instances

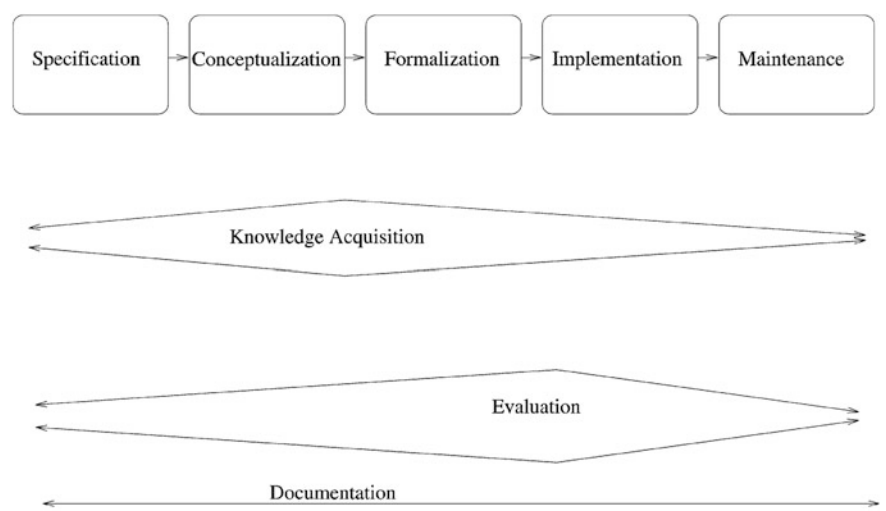


Fig. 2 Activities in ontology development life cycle



Fig. 3 Ontology view using Onto Graf

There are 90 concepts, 25 object properties, and 36 data properties. It will not be possible to cover the whole construction process in the scope of this paper so a view of created ontology is shown below in Fig. 3.

3.1.2 Natural Language Interface

Though there are many ontology languages available, in this work, we have chosen to develop an interface for accessing knowledge from ontology. Our intention is to later on make this system available for real-time use and farmers/naïve users are not aware of any ontology languages. So, we have developed an interface for converting the general natural language questions into standard query language. Some of the key issues that lead to development of such an interface are:

1. Queries natural language are highly ambiguous in nature from linguistics point of view. Efforts are being made to design and develop systems which are best in terms of precision and recall [8].

2. There needs to be a uniform representation mechanism for natural language which is very challenging to achieve. Different systems have different representations and it is extremely difficult to map them into one specific type. Each system has its own vocabulary/schema so it is quite difficult to find out correct mappings which are universally applicable or applicable to a significant number of systems.
3. There are systems developed which have their own knowledge bases and they are able to target specific set of needs. They are not general in nature and it is not possible to adapt to their working as they are trained for specific type of things.

4 Experimental Setup

Keeping in mind the rules proposed by [3] and the various stages involved in the ontology development, the fertilizer ontology is built from scratch. The approach selected is manually-driven. The tool that we are using for the development of FertOnt is Protégé 4.3 (Build 304). Protégé [9] is a free, open source ontology editor and a knowledge acquisition system and is being developed at Stanford University in collaboration with University of Manchester. It supports lot of plug-ins like Pellet reasoner, SPARQL, DL query, etc., which add extra functionalities. It also exports ontology in many formats (RDFS, OWL, etc.). We have used Python programming language to implement the system. The reason behind using Python is the simplicity and powerful programming functionalities provided by the language. Moreover it is free and open source. It is a highly readable language and provides rich support for natural language programming tasks.

1. We have used natural language programming toolkit provided by Python. It provides a very easy to use interface to various corpora and lexical resources. It has vast number of libraries for text processing tasks such as classification, tokenization, stemming, parsing, semantic reasoning, etc.
2. The whole process will begin with the natural language query input by the user. That query will be tokenized into words by making use of the word tokenizer library of nltk. Part of speech tagging gives us the different senses of the words that are used in the input query. After this, the query will be normalized so as to remove the unnecessary portions. At the same time, type of the question will be identified. Once the type of question is identified, it will be matched with the available templates.
3. Then the SPARQL query will be generated. Two natural language questions and their corresponding SPARQL queries are shown in Fig. 4.

The interface part for accessing/retrieving information from ontology is still under development. The basic model is working and it will be further refined.

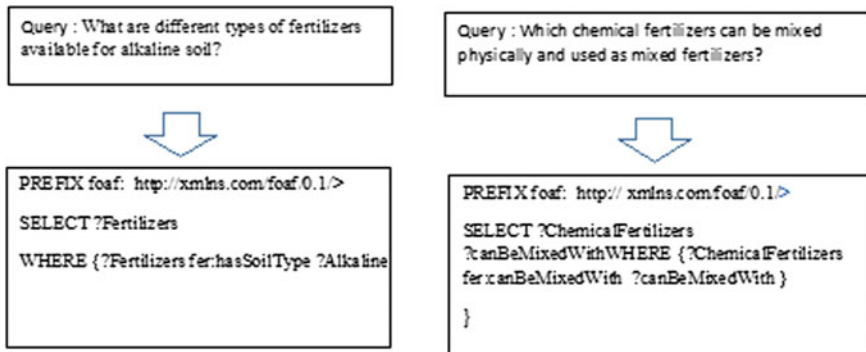


Fig. 4 SPARQL queries for two questions from question set

5 Conclusion

The amount of content that has been generated is increasing day by day. It has become very complex to architect information and knowledge management systems. Ontologies play a key role in building effective knowledge management systems. The paper presents a framework for using ontology as a knowledge base for knowledge management in agriculture domain. Ontology development in the domain of agriculture has been catching a lot of researcher's attention for quite a long time now. However, the subdomain of fertilizer has been poorly explored. It has always been studied in relation to other entities such as soil and crop when it is significantly important to study as separate entity. We have tried to fill this research gap also by developing ontology in the subdomain of fertilizer. Developing an ontology is a time consuming task and requires both manual and expert efforts. Further, this ontology is being created with an intention that in future it may be integrated with soil and crop ontology so that it can be actually used in real-time scenario.

References

1. Sveiby, Karl. The facts about knowledge. In Knowledge management: Cultivating knowledge professionals, edited by Al-Hawamdeh, Suliman. Chandos Publishing, Oxford, 2003. pp. 32–33.
2. Balmissse, Gilles, et al. Technology trends in knowledge management tools. *Int. Journal of Knowledge Management.*, 2007, 3(2), 34–39.
3. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199–220.
4. Liao S., "Knowledge management technologies and applications – literature review from 1995 to 2002", *International journal of expert systems with applications*. (2002). Pp 155–164.
5. Brewster, C., & O'Hara, K. (2007). Knowledge representation with ontologies: Present challenges—future possibilities. *International Journal of Human-Computer Studies*, 65(7), 563–568.
6. Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*.

7. Pinto, H. S., & Martins, J. P. (2004). Ontologies: How can they be built?. *Knowledge and Information Systems*, 6(4), 441–464.
8. C. Wang, M. Xiong, Q. Zhou, Y. Yu, “ PANTO: A portable natural language interface to Ontologies”, *ESWC 2007*, pp. 473–487.
9. <http://protege.stanford.edu/>.

Modularity-Based Community Detection in Fuzzy Granular Social Networks

Nicole Belinda Dillen and Aruna Chakraborty

Abstract Social network analysis is an important task in the modern, globalised world and has several applications in crime, economy, and human psychology. An important aspect of social network analysis is community detection in which groups of closely connected individuals are identified separately from other groups. In this paper, we proposed a new method for detecting communities in a social network. Our method is inspired by fuzzy granular social networks (FGSN) and uses a popular heuristic modularity-based community clustering algorithm. The results obtained from our algorithm correlate well with those obtained by other popular modularity-based detection methods, making it a promising algorithm for community detection in non-overlapping networks.

Keywords Social network analysis · Community detection · Modularity · Fuzzy granular social networks

1 Introduction

A social network is a collection of individuals and their relationships with each other. An example, extremely relevant to our growing internet society would be the online social networks found on websites like Facebook and Twitter. All real-world social networks have one thing in common: proper community structure. Put simply, a community is a collection of individuals that share a common interest. In an online social network like Facebook, community structure could be identified by groups of “friends” with the same “Liked Pages” and, perhaps, even the same “mutual friends”.

N.B. Dillen (✉) · A. Chakraborty
Computer Science and Engineering Department,
St. Thomas' College of Engineering and Technology, Kolkata, India
e-mail: nicolebdillen@gmail.com

A. Chakraborty
e-mail: aruna.stcet@gmail.com

Based on the facts mentioned above, one could see how detecting community structure in networks could help in studying or predicting the overall behaviour of the networks. It could help business organisations target sections of society that would most likely respond positively to their products. It could help law enforcement agencies in unfolding pockets of criminal organisations which may not have been initially prominent. In large networks, it could even aid in the detection of important “key” personalities such as influential politicians or eminent scientists.

Since the mid-90s, several community detection algorithms have been developed. One of the most popular and earliest of these was the Girvan-Newman algorithm [1] which partitions the graph into a number of communities by removing connections that are most likely to occur “between” communities. Another algorithm by Pons and Latapy [2] used random walks to detect communities. An approach by Newman [3] used a fast modularity optimisation algorithm which was later improved by Clauset, Newman and Moore (the CNM algorithm) [4]. Wakita and Tsurumi [5] developed a new metric known as *consolidation ratio* which attempts to balance the communities detected by the CNM algorithm. Both Newman’s original algorithm as well as the CNM method inspired the well-known Louvain algorithm [6] which uses a greedy modular optimisation technique to accomplish the community detection task.

Many of the newer community detection algorithms include an additional scope of detecting overlapping communities, i.e., communities whose nodes may belong to other communities as well. Work in this area includes a modified modularity optimisation algorithm [7] for detecting overlapping communities and an algorithm based on the concept of Fuzzy Rough set theory [8].

In our paper, we have proposed a novel algorithm that is applicable to real-world social networks without any overlapping communities. Our concept is inspired by Fuzzy Granular Social Networks [9] as well as the Louvain algorithm [6]. While the Louvain algorithm deals with distinct, individual nodes, our algorithm extends this principle to the domain of fuzzy granular social networks (FGSN), the main aim of which is to reduce the set of nodes to a smaller set of granules.

Our paper is organised as follows: Sect. 2 deals with the preliminary information related to our algorithm while Sect. 3 describes the said algorithm. The results obtained are discussed in Sect. 4. Finally, Sect. 5 deals with our conclusions and possible future work.

2 Preliminary Concepts

In this section, we formally define and explain some of the necessary concepts related to social network analysis which have been used by our proposed algorithm.

Definition 1 Formally, a social network is a graph $G(V, E)$ where:

1. V is the set of all vertices or nodes or individuals
2. E is the set of all edges or links that interconnect the vertices in V

Definition 2 A community is a subset of nodes that are densely interconnected while being sparsely connected to the other nodes of other communities.

Our objective is, therefore, to identify such groups of densely interconnected nodes and classify them as communities. To do this, we model our social network system in the fuzzy domain using the concept of granules which we describe below.

Most algorithms take into account each and every node of a network. However, we could drastically reduce the number of computations if we group “similar” nodes together to form what are known as “granules” and perform the remaining computations solely on these granules. In fact, this is the very same approach used in creating FGSN [9]. The next four definitions serve to clarify this concept.

Definition 3 A granule [10] is a collection of similar, indistinguishable objects that can be treated as an independent unit. In the context of social networks, a granule, denoted by A_c , is represented by a centre vertex c , and a node’s relationship (usually a distance function) with c defining its membership in the granule.

Realistically speaking, some nodes may have equal or different memberships in multiple granules instead of just one. To account for this, the fuzzy domain [11] has been incorporated in the membership assignment of nodes to various granules. We denote this fuzzy membership as $\mu_c(v)$ which denotes the membership of node v (a monotonically non-increasing function) in the granule represented by centre c .

Definition 4 According to the FGSN [9] theory, the membership of a node v in a granule represented by centre c is denoted by $\mu_c(v)$, and defined in Eq. (1) below:

$$\mu_c(v) = \begin{cases} 0 & \text{for } d(c, v) > r \\ \frac{1}{1 + d(c, v)} & \text{otherwise} \end{cases} \quad (1)$$

where $d(c, v)$ is the distance between node v and granule centre c , and r is the desired granular radius which may be varied. When the distance is 0, i.e., the vertex is c itself, $\mu_c(v) = 1$ while $\mu_c(v) = 0$ for infinite distance. Also, these membership values must be normalised as a node may belong to a number of granules with varying membership. The normalised membership value is then,

$$\tilde{\mu}_c(v) = \frac{\mu_c(v)}{\sum_{i \in C} \mu_i(v)}. \quad (2)$$

Thus, we have:

1. C , the set of vertices each representing a particular granule, and
2. $Gr = \{A_c | \forall c \in C, \sum_{v \in V} \tilde{\mu}_c(v)/v\}$, the set of all granules.

Another important aspect of social networks is embeddedness.

Definition 5 The embeddedness for a pair of granules, centred at a and b respectively, is the extent to which one is embedded in the other. It is nothing but the cardinality of the intersection of both granules and is denoted by $\varepsilon(a, b)$:

$$\varepsilon(a, b) = |A_a \cap A_b| = \sum_{v \in V} \min(\tilde{\mu}_a(v), \tilde{\mu}_b(v)). \quad (3)$$

We now provide a brief background of the Louvain algorithm [6] that is used to detect communities in social networks: the Louvain algorithm is a greedy optimisation algorithm which works on the principle of modularity. Like other optimisation algorithms [3, 4], its objective is to maximise the modularity by placing nodes in communities that result in a local maxima.

Definition 6 Modularity Q is a measure used to provide a qualitative assessment of the community partitions that have been detected in the network. It conveys the difference between the actual density of interconnections between nodes in a detected community and the corresponding connections in a random network possessing the same degree distribution as that of the actual network [3]:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j). \quad (4)$$

where A_{ij} is the weight of the edge between vertices i and j , k_i is the total weight of all edges linked to i , c_i is the community to which i is assigned and $\delta(c_i, c_j)$ is 1 when both i and j belong to the same community and is 0 otherwise. The total weight of all edges in the network is m , where $m = \frac{1}{2} \sum_{i,j} A_{ij}$.

The Louvain algorithm consists of two stages. The first, also called the “iterative stage” is the greedy stage which iteratively looks for the local maxima of the modularity. Each node is initially considered a single community. For each node i , we compute the change in modularity obtained by removing i from its community and placing it in the community of one of its neighbours. The node i is placed in the community for which this modularity change is both positive and maximum. This process is repeated for all other nodes in the network. The entire stage is then repeated iteratively until no further increase in modularity can be obtained.

The second stage, or “coarse-graining” stage, groups all the nodes belonging to a community into a single unit. A new network is formed whose nodes correspond to the communities detected during the iterative stage. Here, a link between two nodes is simply the sum of weights of the connections between the nodes of the corresponding communities. Similarly, self-loops may also be generated which have weights equal to the sum of intra-connections between nodes of the same community. The first stage is then reapplied to the new adjacency matrix. The two stages are repeated until the modularity cannot be optimised any further. A hierarchy is thus obtained consisting of the communities detected after each phase of iteration and coarse-graining.

3 Proposed Algorithm

In brief, our algorithm is simple. We first choose certain nodes as granule centres. This is done by computing the average degree of all nodes in the network and choosing only those nodes with a degree greater than the average as granule centres. We set the network diameter as the granule radius and construct the set of all granules according to the methods described in the previous section. Next, we seek to construct a new “network” whose nodes correspond to the granules. We assign a link between two nodes in our new network whose weight is equal to the embeddedness between the corresponding two granules. Note that this is applicable to self-loops as well. In the case of a self-loop, the loop weight will simply be the cardinality of the granule itself.

After we construct the adjacency matrix for our new network, we detect “granular communities” in it by means of the Louvain algorithm. We use these granular communities to construct the actual set of corresponding communities for the vertices in the social network. We first construct a Fuzzy-Rough community matrix [9] in the following manner: for every granular community g_i , we construct a corresponding community C_i in which a vertex v 's membership to C_i is set to 1 if all its positive granular memberships involve only those granules that have been assigned to g_i . If v is assigned positive memberships in granules belonging to multiple granular communities, its membership to C_i is equal to sum of its memberships of all granules in g_i . Obviously, if v possesses 0 membership in all granules of g_i , it will be assigned a membership of 0 to C_i . Finally, for every vertex v , we look for the community in which v has the highest membership value and set this value to 1. All other membership values are set to 0. Our algorithm, which we now call “GranLouv” is formally stated below.

Algorithm: GranLouv

1. *Start*
2. Set granule representative set: $C \leftarrow \{v: \text{degree}(v) > \text{average degree}\}$.
3. Form granule set $Gr = \{A_c \mid \forall c \in C, \sum_{v \in V} \tilde{\mu}_c(v) / v\}$ using distance function (1) and membership function (2).
4. Consider each granule as a ‘vertex’ and construct an adjacency matrix as below:
 - a. initialise an $N \times N$ matrix M where N is the number of granules.
 - b. $M(i, j) \leftarrow \varepsilon(a, b)$, from equation (3).
5. Granular-communities $g \leftarrow \text{Louvain}(M)$.
6. Obtain Fuzzy Rough Community matrix from g .
7. $\forall v \in V$, find C_i for which $C_i(v)$ is maximum. Set $C_i(v)=1$ and $C_j(v)=0, \forall j \neq i$
8. *End*

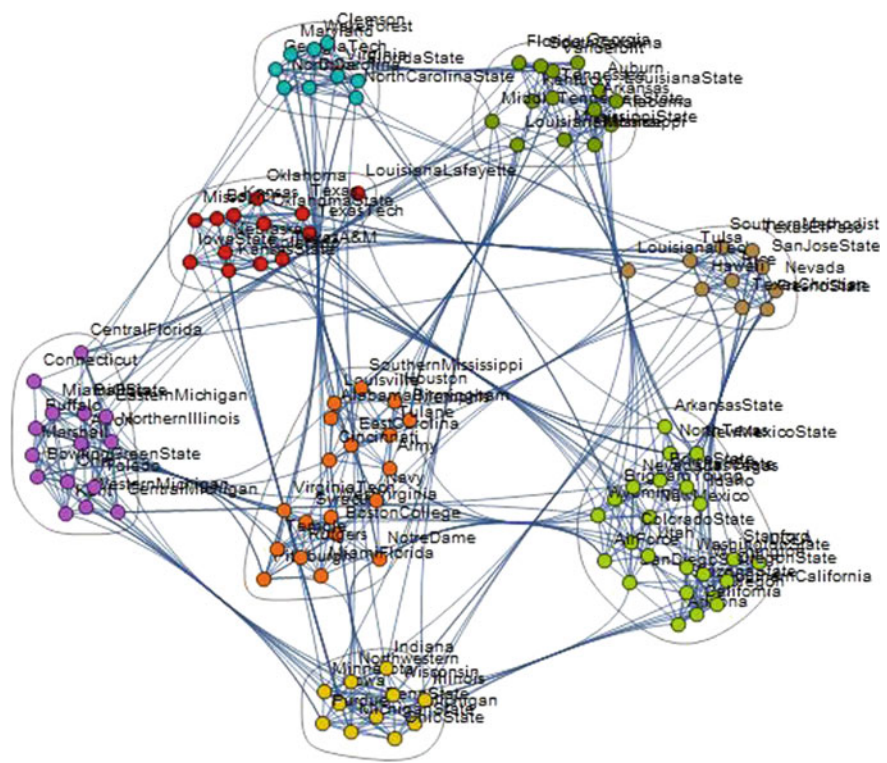


Fig. 3 American college football social network: communities detected using GranLouv

4 Application and Results

Our algorithm was tested using a 2.40 GHz dual core CPU with 2.00 GB RAM and was implemented with Mathematica 10.0. We considered three different real-world datasets, namely, the Dolphin [12], Les Miserables [13] and the American College Football [1] social networks. The results obtained for each of these networks compared with those obtained by various reference algorithms are provided below (Figs. 1, 2, 3 and Tables 1, 2 and 3). In all cases, we considered the highest modularity obtained in a series of iterations of the algorithm.

Table 1 Dolphin social network: modularity

Serial no.	Algorithm	Modularity
1.	Louvain	0.518
2.	GranLouv	0.509

Table 2 Les miserables social network: modularity

Serial no.	Algorithm	Modularity
1.	Newman and Girvan [14]	0.540
2.	Louvain	0.555
3.	GranLouv	0.534

Table 3 American college football social network: modularity

Serial no.	Algorithm	Modularity
1.	Girvan-Newman [1]	0.601
2.	Louvain	0.604
3.	GranLouv	0.599

5 Conclusion and Future Work

We have seen from our results that our algorithm produces results comparable to other popular algorithms. While other modularity-based algorithms have considered each and every node of the network during the detection process, we have accomplished the optimisation task by considering only a few granules instead.

Scope for improvement lies in the selection of granule centres as, in our implementation, we have considered only those nodes with degree greater than the average degree of the network. However, this may not be the best way to select the most significant or important nodes. A better selection algorithm could yield even better granular communities and, thus, better final communities. Our future work will include modifying the algorithm to address the granule centre selection problem mentioned above as well as extending our algorithm to accommodate overlapping communities.

Acknowledgments The authors of this paper wish to convey their thanks to Prof. Sankar Kumar Pal as well as Mr. S. Kundu from the Center for Soft Computing Research, Indian Statistical Institute, for their continued support and assistance provided during the conception of this paper.

References

1. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA, 99(12):7821–7826, June 2002.

2. Pons, P., Latapy, M., 2006 Journal of Graph Algorithms and Applications 10 191–218.

3. Newman M.E.J.: Fast algorithm for detecting community structure in networks. Phys. Rev. E 69, 066133 (2004).

4. Clauset, A., Newman, M., Moore, C.: Finding community structure in very large networks. Physical Review E, 70(6):1–6, 2004.

5. Wakita, K., Tsurumi T.: 2007 Proceedings of IADIS international conference on WWW/Internet 2007 153.

6. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008 (12 pp) doi:[10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008). arXiv:<http://arxiv.org/abs/0803.0476>.

7. Ganganathy, N., Chenz, G., Cheng, C.T.: Detecting Hierarchical and Overlapping Community Structures in Networks. 2014 International Symposium on Nonlinear Theory and its Applications, NOLTA2014, Luzern, Switzerland, September 14–18, 2014.

8. Kundu, S., Pal, S.K.: FGSN: Fuzzy Granular Social Networks – Model and applications. Information Sciences, vol. 314, 1 September 2015, pp. 100–117.

9. Kundu, S., Pal, S.K.: Fuzzy-rough community in social networks. Pattern Recognition Letters (2015), <http://dx.doi.org/10.1016/j.patrec.2015.02.005>.

10. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* 90 (1997) 111–127.
11. Zadeh, L.: Fuzzy sets, *Inform. Control* 8 (1965) 338–353.
12. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.: The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396–405, 2003.
13. Knuth, D.E.: *The Stanford GraphBase: A Platform for Combinatorial Computing*. Addison-Wesley, 1993.
14. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Phys. Rev. E*, vol. 69, p. 026113, 2004.

Auto-Characterization of Learning Materials: An Adaptive Approach to Personalized Learning Material Recommendation

Jyoti Pareek and Maitri Jhaveri

Abstract The need of a learning platform where individual learner deserves his/her own learning path towards mastering a subject is vigorously increasing. This self-directed and adaptive learning enforces the personalised learning environment to adapt to the needs and learning style of learner. We propose to model the multidimensional characteristics of the learning material and the knowledge acquisition pattern of learner for personalized recommendations. This model recommends learning materials to the learner whose characteristics match with those learning materials, which have benefited the learner most in past. Post-study cognitive knowledge is tested for establishing the benefits to learner. The system automatically generates and evaluates compare-and-contrast questions presented to the learner. Satisfactory results are obtained in automatic annotation of learning materials and performance evaluation score generation. F1 score of 0.8404 and 0.650 was, respectively, obtained while evaluating the identification of learning material attributes and generation of compare-and-contrast questions.

Keywords Personalized learning · Learning material metadata · Cognitive knowledge · Compare-and-contrast questions

1 Introduction

The current education scenario is shifting from the classic classroom “one-size-that-fits-all” type of learning to self-directed, student-centric and exemplary learning. Personalized learning pedagogy facilitates each learner to choose his/her own learning resources. A tool in a personalized learning environment is

Jyoti Pareek (✉)

Department of Computer Science, Gujarat University, Ahmedabad, India
e-mail: drjyotipareek@yahoo.com

Maitri Jhaveri

Banasthali University, Banasthali, Rajasthan, India
e-mail: maitrijhaveri@yahoo.com

required, which supports every student in locating the most suited learning material according to his/her learning style. The preferred learning style guides the way a student learns. It may be one or combination of visual, aural, verbal, physical, logical, social and solitary learning styles. If a student is able to score well after reading a material then we can say that the corresponding learning style suits him/her. Once such a material is found by which the student has benefited the most, then recommending similar type of learning material to him/her will definitely improve his/her performance in the overall subject. Similarity between the learning materials can be established by comparing their attributes. Attributes to the learning materials can be assigned on the basis of learning object metadata. Various standards are present for learning object metadata such as IEEE LOM, SCORM, Dublin Core, IMS LRM, CanCore and UK LOM Core. Learning material being a technical document becomes very important that some learners may require the explanation of complex technical ideas in simple terms. Some may require ease in the language used where scientific and technical information is presented in a clear and easier understanding way. Some learners may require that data should be concise and language should be straight forward. The clarity and readability of a document can be analyzed by knowing the amount of active words, assertive sentences, long sentences, long words and long paragraphs. The ease of understanding can be analyzed by the number of examples, case studies and figures present in the document. Hence we propose to characterize these learning materials by ease of language used, explanation of concepts through usage of examples, figures, tables, case studies and usage of assertive statements and active voice. Hence we propose to quantify the aforementioned characteristics of learning material. To categorize the learning materials, we propose the generation of four new attributes such as clarity score, readability score, understandability score and average performance score.

2 Literature Review

Learning object repositories developed world wide maintain the metadata of each learning resource. Annotation of each learning resources leads to its reusability. The resource can be classified based on its metadata and hence can cater to the personalised requirement of each learner. Ghauth and Abdullah [1] propose recommendation of learning material on bases of content similarity and good learner's rating. Peer learning and social learning theories are used for recommendations. Zhong and Li [2] have proposed solution of mapping collaborative filtering problem to text analysis problem using combination of implicit and explicit features of learners and resources. They claim improved and accurate results compared to memory-based techniques. Recommendation of learning resources using collaborative filtering was proposed by Wan et al. [3]. They tried to solve the problem of absence of face-to-face communication with the teacher. They record learning behaviour as a sequence of events and apply sequential pattern mining on them. Garcia et al. [4] proposes a collaborative data mining tool which works on mining of association rules. Li et al. [5] proposes

mining of web logs for integrated collaborative filtering and sequential pattern mining recommending learning resources to individual learner. It also incorporates the learning materials read by the learner. A framework for learning material recommender system was proposed by Salehi et al. [6] which keeps track of learner's interest. It models multidimensional attributes of each learning material. An approach for recommending learning materials was proposed by Salehi and Nakhai Kamalabadi [7]. It focused on studying the sequential pattern of learning materials read by the learner. Association rules were then applied on the patterns studied. Learner's preferences were modeled through a compact tree. Another hybrid recommendation approach for recommendation of learning material was proposed by Salehi et al. [8] for improvement in quality and accuracy of results. It was based on genetic algorithm and multidimensional user preference model. A new similarity measure was incorporated and nearest neighbourhood algorithms were used. Ley et al. [9] have worked on effect of personalized scaffolding in the learning process. Shaw et al. [10] propose frameworks for modules like the content map, learning nuggets and recommendation algorithms. Tasi et al. [11] propose recommendation of SCORM-compliant learning objects. These objects lie in internet repositories. The degree of relevance of learning objects with respect to learner's preference is measured based on preference-based and correlation-based approaches. Chen et al. [12] propose to combine collaborative filtering with learning material response theory to recommend personalized path to learners. Kay [13] presents a model for the lifelong user as a first class citizen, existing independently of any single application and controlled by the learner. Chen et al. [14] use fuzzy item response theory to recommend personalized courseware. It also claims to estimate learner's ability and courseware difficulty. An adaptive ranking mechanism is proposed by Tsai et al. [11] to measure the relevance of the learning resource with respect to learner's preferences and interests of his/her neighbors. Yu et al. [15] emphasizes on context aware learning. It uses ontology to semantically model the knowledge about learner, content and the domain under study. Their system reduces the time taken to search personalized and relevant learning objects. A personalized learning system was proposed by Baylari and Montazer [16] to recommend personalized learning materials through artificial neural networks. It also works on item response theory. Imran et al. [17] proposed a framework for content recommendation on a given subject. It uses vector space model. It tracks good learners' rating and content in learning resources. Khribi et al. [18] proposed a two module approach to provide online recommendations to new learners based on existing learners' navigation history. It does not require learner's explicit feedback. It claims to recommend learning resources based on learner's needs and goals. Romero et al. [19] combined web mining with the AHA system for providing personalised recommendations. None of this work emphasizes much on the characteristics of the learning materials. In studying sequential patterns of learning materials only the type of learning material such as case study, exercise, questionnaire, etc., are considered. The rating given by good student is incorporated but the category of material in which learner has performed well is not taken into account. For example, some learners may perform well if the learning material is simple to read. Some may perform well if additional learning tools other than plain text are used such

as examples, tables, graphs, figures, case studies, etc. To our knowledge, methods proposed till now for personalised learning material recommendation do not maintain the record of attributes of learning material in which the learner has performed well. We propose to identify learner’s performance in a learning material rather than just taking his/her ratings. We further propose to automatically categorize all learning materials by its simplicity, readability and understand ability so that we can identify the category of learning material in which the learner tends to perform well.

3 Proposed Work

Figure 1 shows the architectural view of the model proposed. It is a recursive procedure where learner’s feedback is taken into account for further recommendations. It is a two step procedure where in the first step the preference of the learner is decided by evaluating his/her past performance. The second step does the personalised recommendation by matching the attributes of all candidate learning materials for the topic under study and those of the most preferred material.

All important functionalities are explained as under.

3.1 Generation of Learning Material Characteristics

3.1.1 Readability Score

Apache POI API provides the interface to fetch paragraphs present in the document. According to the basic rules of grammar, the paragraphs are split into sentences and sentences are split into words. Readability score is established as

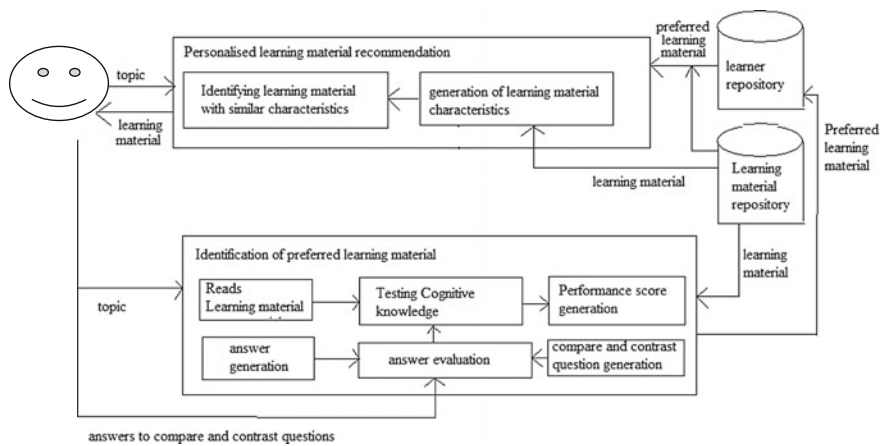


Fig. 1 Architectural diagram of the proposed model

$$((\# \text{ long paragraphs} \div \# \text{ paragraphs}) + (\# \text{ long sentences} \div \# \text{ sentences}) + (\# \text{ long words} \div \# \text{ words})) * 100$$

Our survey on articles of good technical writing document identifies that, a paragraph is considered as long if it contains more than six sentences. A sentence is considered to be long if it contains more than 30 words. A word is said to be long if it contains more than 20 characters.

3.1.2 Clarity Score

A sentence is said to assertive if it is none of question type sentence, order type, or exclamatory type. Assertive sentences and the words in active voice have energy and directness. They motivate the reader turning the pages. The active voice offers many advantages to the technical writer. It emphasizes the person to perform the action, allows the reader to grasp the main idea easily and employs simple sentence structure.

Clarity score is established as

$$((\# \text{ assertive sentences} \div \# \text{ sentences}) + (\# \text{ active words} \div \# \text{ words})) * 100$$

3.1.3 Score of Understanding

It is established as total number of examples and figures. The API provides the feature to extract figures from the Word Document. But extracting examples from the document requires special efforts. We have used a pattern-based mining approach to extract them. We have studied multiple documents containing examples. We have analyzed that all examples do follow certain patterns. We have made a list of such patterns. Table 1 shows the list of patterns considered in extracting examples.

Table 1 Templates for example extraction

For instance	An example of this is	Such as	Like	Typical example	Provides an example of
Classy	It is a very good model of	Set a good example	It is a model of	Models	Typical example
And	Specimen	Illustrate	e.g.	Lemma	i.e.
For e.g.	Example	For example	Such as	By way of illustration	In particular
As an illustration	Typical case	Case in point	Paradigm	Sample	

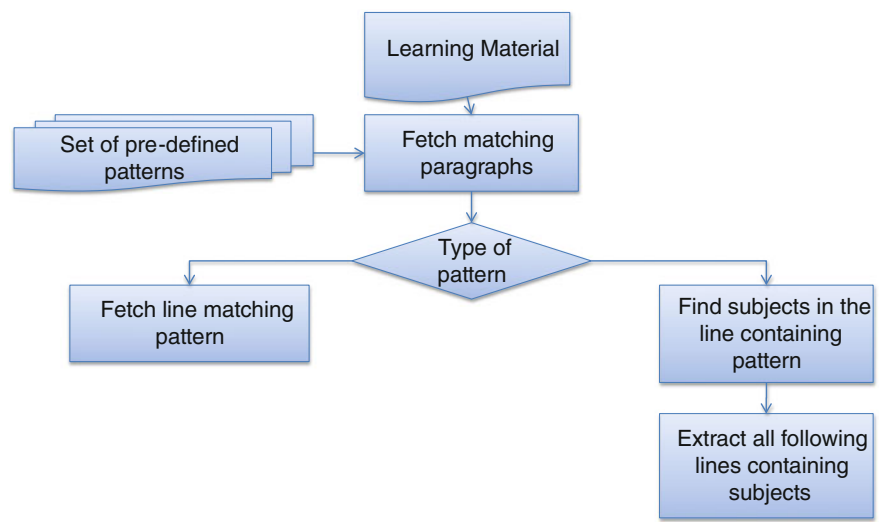


Fig. 2 Extraction of single and multiline attributes

We manually analyzed the document to find out the examples with these patterns. Then we analyzed these examples and found out that some patterns mostly yields examples that only last for a single sentence and some last for more than one sentence. So we classified them in two types, one that forms examples of single-line and one that form examples of multi-lines. Stanford deterministic conference resolution system is used for finding sentences which have indirect references to subject. Figure 2 explains the extraction of single-line and multiline examples.

3.1.4 Average Performance Score

Average performance score of learning material is obtained by averaging the performance score of each learner who has read the corresponding learning material. For analyzing the performance of the learner, we present a list of questions to the learner to test his/her knowledge. Currently, our system concentrates on testing the cognitive knowledge. The learner is provided with a list of compare-and-contrast questions. These questions are autogenerated by the system. The answers provided by the students are autoevaluated by the system. The evaluation is done against the answers generated by the system itself.

Autogeneration of Compare-and-Contrast Questions

We propose to generate compare-and-contrast questions from a learning material through pattern-based mining. We have chosen the domain of “operating systems in

Table 2 List of templates for compare-and-contrast entity extraction

Difference between/in	Like	And	Either-or
Both-and	Neither-nor	Similar-to	Rather-than
While	Although	Whereas	

computer science” to evaluate our approach. We have studied various documents of this subject to analyze and identify patterns which are normally followed to give us concepts which can be compared. Table 2 gives us the list of such patterns.

Automatic Answer Evaluation

In answer extraction, we have extracted all the sentences that talk about the entities. At a time we find sentences or information on one entity. To find the information for an entity, we analyze each sentence of the document. For analyzing, we have used Stanford Parser and Co-reference Resolution API. Using these APIs, we find the sentences which talk about the entity. Cosine similarity function is used to match the answer given by the learner and system generated answer. The similarity score for all the answers given by the learner is generated against the corresponding system generated answers say $s_1, s_2, s_3 \dots s_n$, where n is number of compare-and-contrast questions pertaining to document d. The performance score for each learner ‘i’ for document d is then generated as $P_{id} = (s_1 + s_2 + \dots + s_n)/n$. Performance score of document d is given by $P_d = (p_{1d} + p_{2d} + \dots + P_{md})/m$, where m is the total number of learners who have read the document d.

The scores of readability, clarity and understand ability of all learning materials belonging to the topic of study are matched with that learning material which the learner has already read and performed best. All those matching learning materials whose average performance score is high are then recommended to the learner.

4 Experimental Results

To test the effectiveness of our proposed model, we have developed a prototype system. The prototype is developed in java. Pattern mining approach is used for autogeneration of learning material characteristics. System contains a repository of learning materials. All the learning materials are Microsoft Word Documents. The extensions supported are .doc and .docx. Microsoft documents cannot be accessed or read using java.io package. Therefore, system uses Apache POI jar to suit our purpose. Java APIs for manipulating various file formats based upon the Microsoft’s OLE 2 Compound Document format (OLE2) and Office Open XML standards (OOXML) is provided by Apache POI. Apache POI has a complete API for porting other OOXML and OLE2 formats and welcomes others to participate. To identify patterns we have used Stanford parser and Stanford log-linear part of



Fig. 3 Sample output showing the instructional content and questions to be answered from a sample learning material

speech tagger. We have used Stanford deterministic conference resolution system for finding sentences which have indirect references to subject. This system implements the multi-pass sieve conference resolution system also known as anaphora resolution system. The topic under study is “Memory management in operating system”. Currently we have performed testing on 15 documents total of capacity 6 MB and 350 pages. Figure 3 shows the output screen showing the instructional content and questions to be answered from a sample learning material.

5 Evaluation

The efficiency of our working prototype can be evaluated in three parts.

1. Efficiency in Identification of Learning Material Characteristics.

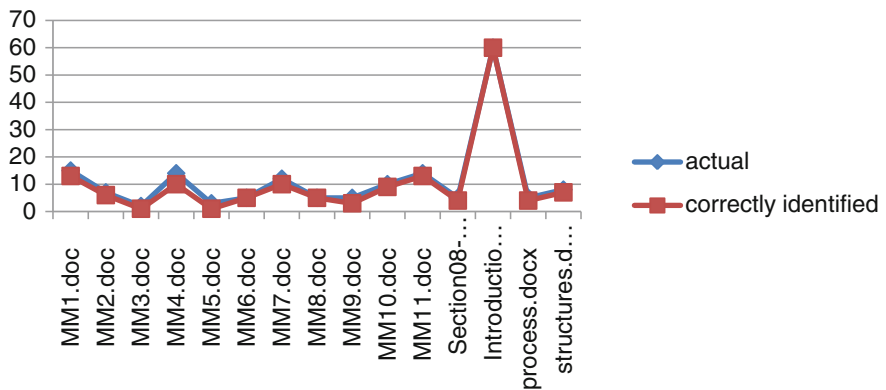
Graph 1 shows the number of correctly identified examples for each document versus actual number of examples for each document with average precision as 0.884 and average recall as 0.801. A higher precision and recall is obtained in identification of paragraphs, assertive sentences and active words.

2. Efficiency in Compare-and-Contrast Question Generation.

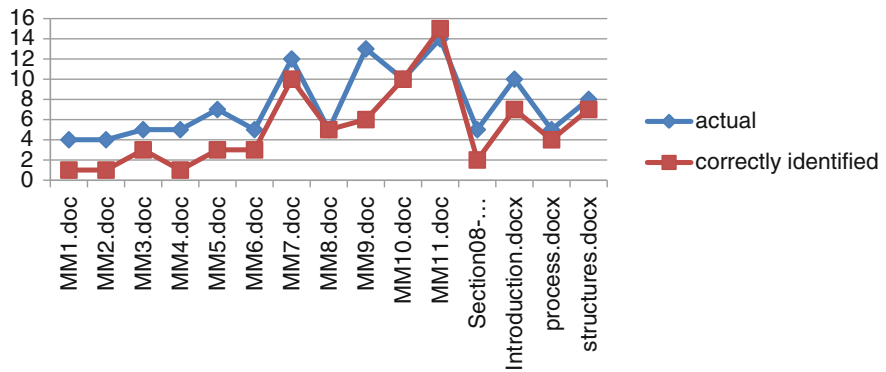
Graph 2 shows the number of correctly identified compare-and-contrast questions for each document versus actual number of compare-and-contrast questions for each document with average precision as 0.689 and average recall as 0.631.

3. Efficiency in Personalised Learning Material Recommendation

Testing of this application is ongoing right now. We have tested this application with few students as of now. To start with we have assigned performance scores to the learning materials as per expert’s advice. This score will be updated every time



Graph 1 Statistics of example extraction



Graph 2 Statistics of compare-and-contrast entities extraction

Table 3 Learner's performance score of three learners who have read three learning materials

Documents	Learner	Learner performance score (%)
Introduction.doc	1	62
Introduction.doc	2	30
Introduction.doc	3	45
Process.docx	1	40
Process.docx	2	67
Structures.doc	1	45

a student reads the learning material and appears in test. Table 3 currently depicts the learner's performance score of three learners who have read three learning materials.

Table 4 System-generated scores of the learning materials present in the repository

Document	Clarity Score (%)	Readability score (%)	Number of figures	Number of examples	Average performance score (%)
MM1.doc	50	72	7	13	56
MM2.doc	57	23	0	6	34
MM3.doc	60	30	4	1	33
MM4.doc	54	20	24	12	45
MM5.doc	55	6	30	1	30
MM6.doc	25	33	0	6	34
MM7.doc	50	26	4	16	40
MM8.doc	48	4	16	6	32
MM8.doc	45	5	5	3	20
MM10.doc	32	7	1	10	34
MM11.doc	47	26	11	15	45
Section 08-Memory_ Management.doc	30	1	10	5	23
Introduction.doc	53	60	14	60	46
Process.doc	60	34	40	6	56
Structures.doc	50	42	41	7	50

Table 4 gives the scores of the learning materials present in the repository. Learner 1 who has scored 62 % by reading introduction.doc is recommended three materials namely doc 2.doc, process.doc and structures.doc. At least two characteristics match with the material in which learner 1 has scored highest. This is the sample output generated and exhaustive testing is still going on.

6 Conclusion and Future Work

This paper annotates each learning material with the metadata such as readability score, clarity score and score of understanding. These characteristics are very important when a novice learner wants to learn from a technical document. To achieve personalization, our model analyzes and records the performance pattern of the learner in the previously read learning materials. Once the learning material which has benefited the learner most is found then learning materials with similar metadata is recommended. The post-study performance of the learner is evaluated by making him to take the system generated test. This test currently checks the cognitive knowledge through compare-and-contrast questions. The test paper generation and its evaluation are fully automated. Exhaustive testing is our next task with learning materials in pdf as well as power point formats. To improve the accuracy in learner's performance assessment, more categories of open-ended questions will be included in our future work. In order to improve accuracy in similarity identification among learning materials, we further plan to extract more metadata from learning materials. We also plan to model the relationships among the concepts from different learning materials.

References

1. Ghauth, K.I., Abdullah, N.A.: Learning Materials Recommendation Using Good Learners' Ratings and content-Based Filtering. *Educational Technology Research and Development*, 58, 6, 711–727 (2010).
2. Zhong, J., Li, X.: Unified collaborative filtering model based on combination of latent features. *Expert Systems with Applications*, 37, 8, 5666–5672 (2010).
3. Wan, A.T., Sadiq, S. and Xue Li: On Improving Learning Outcomes through Sharing of Learning Experiences. In: 10th IEEE International Conference on Advanced learning Technologies. Sousse, Tunisia, July pp. 5–7. IEEE Computer Society: California. (2010).
4. Garcia, E., Romero, C., Ventura, S.: Collaborative Educational Association Rule Mining Tool. *Internet and Higher Education*, 14, 2, 77–88 (2011).
5. Li, Y., Niu, Z., Chen, W. and Zhang, W.: Combining Collaborative Filtering and Sequential Pattern Mining for Recommendation in ELearning Environment. In: 10th International Advances in Web-Based Learning, 305–313 (2011).
6. Salehi, M., Nakhai Kamalabadi, and Ghaznavi Ghouschi, M.B Personalized Recommendation of Learning Material Using Sequential Pattern Mining and Attribute Based Collaborative Filtering, *Education and Information Technologies*, 17, 4 1–23 (2012).

7. Salehi, M and Nakhai Kamalabadi, I.: Hybrid Recommendation Approach for Learning Material Based on Sequential Pattern of the Accessed Material and the Learner's Preference Tree, *Knowledge-Based Systems*, 48, 57–69 (2013).
8. Salehi, M., Pourzaferani, M. and Razavi, S.A.: Hybrid Attribute-Based Recommender System for Learning Material Using GeneticAlgorithm and a Multidimensional Information Model. *Egyptian Informatics Journal*, 14, 1, 1–23 (2013).
9. Tobias Ley, Barbara Kump, Cornelia Gerdenitsch: Scaffolding Self-directed Learning with Personalized Learning Goal Recommendations. *User Modeling, Adaptation, and Personalization*, Lecture Notes in Computer Science Volume 6075, 2010, pp 75–86.
10. Shaw, C., Larson, R. and Sibdari, S.: An Asynchronous, Personalized Learning Platform—Guided Learning Pathways (GLP). *Creative Education*, 5, 1189–1204. (2014).
11. Tsai, K.H., Chiu, T.K., Lee, M.C. and Wang, T.I.A.: Learning Objects Recommendation Model Based on the Preference and Ontological Approaches In: 6th IEEE International Conference on Advanced Learning Technologies, Kerkraide, July 5–7, 2006.
12. Chih-Ming Chen, Hahn-Ming Lee, Ya-Hui Chen.: Personalized e-learning system using Item Response Theory, *Computers and Education*, 44, 3, 237–255 (2005).
13. J. Kay.: Lifelong Learner Modeling for Lifelong Personalized Pervasive Learning, 1, 4, 215–228 (2008).
14. Chih-Ming Chen, Ling-Jiun Duh and Chao-Yu Liu.: A Personalized Courseware Recommendation System Based on Fuzzy Item Response Theory In: IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04). Taipei, Taiwan, March 28–31, pp 305–308, 2004.
15. Zhiwen Yu, Yuichi Nakamura, Seie Jang, Shoji Kajita, and Kenji Mase.: Ontology-Based Semantic recommendation for Context Aware E-Learning. In: International Conference on Ubiquitous Intelligence and Computing, Hong Kong, China, July 11–13, pp 898–907, 2007, Springer Berlin Heidelberg.
16. Ahmad Baylari and Montazer, Gh.A.: Design a personalized e-learning system based on item response theory and artificial neural network approach. *Expert Systems with Applications*, 36, 4, 8013–8021 (2009).
17. Khairil Imran, Bin Ghauth and Nor Aniza Abdullah.: Building an E-Learning Recommender System using vector Space Model and Good Learners Average Rating–Vector space model. In: 9th IEEE International conference on Advanced Learning Technologies (ICALT 2009). Riga, Latvia, July 15–17, 2009, IEEE Computer Society: California.
18. Khribi, M.K, Jemni, M. and Nasraoui, O.: Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In: 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), Santander, Cantabria, July 1–5, 2008. IEEE Computer Society: California.
19. Romero, C., Ventura, S., Zafra, A. and de Bra.P.: Applying Web Usage Mining for Personalizing hyperlinks in Web-Based Adaptive Educational Systems. *Computers and Education*, 53, 3, 820–840 (2009).

Hadoop with Intuitionistic Fuzzy C-Means for Clustering in Big Data

B.K. Tripathy, Dishant Mittal and Deepthi P. Hudedagaddi

Abstract In recent days, industry and academia have been trying to address the data handling issues with respect to big data. This has led to development of new computing arenas in the fields of data mining and analysis of data which are the need of the hour. One of the techniques to handle large data is by making clusters of the similar data. But this technique is complex as well. This paper proposes a new algorithm/technique of data clustering where Intuitionistic Fuzzy C-Means (IFCM) is used along with Hadoop to produce high-quality clusters and thereby making clustering on very large data more efficient. The results of the proposed algorithm are demonstrated with the help of UCI data sets. Performance metrics like Accuracy, SSW, SSB, DB, DD, and SC indices are used for comparison of the obtained results with Parallel K-means (PKM) and modified Parallel K-means (MPKM).

Keywords Clustering · Hadoop · Big data · Intuitionistic · FCM · IFCM

1 Introduction

For many years data collection was a problem, but now scientists and researchers are facing challenges in processing this accumulating avalanche of data. This has raised the need for powerful tools for handling the ever-increasing big data.

B.K. Tripathy (✉) · D.P. Hudedagaddi
SCSE, VIT, Vellore 632014, Tamil Nadu, India
e-mail: tripathybk@vit.ac.in

D.P. Hudedagaddi
e-mail: deepthiph@gmail.com

Dishant Mittal
Johnson Controls, Mumbai 400042, Maharashtra, India
e-mail: dishantmittal31@gmail.com

Clustering is one such major tool. A good clustering algorithm is expected to produce quality clusters by having elements similar to one another within the cluster and dissimilar to the ones in external clusters. This ‘binning’ of data helps in providing visualisation of large amount of data and hence makes it more readable. The existing and most commonly used K-means and fuzzy C-Means are the notable techniques for creation of clusters. But many of these traditional methods fail in execution time, speed, accuracy, and hence provide poor results on directly using them on big data. Hence an iterative process is proposed here. Hadoop cluster is used by IFCM to provide high-quality clusters.

A cluster in which large piles of unstructured data are stored and analysed in different environments of distributed computing can be called as a Hadoop cluster. These kinds of clusters execute on the distributed processing software which is a open source on less expensive, simple computers [1]. These are extensively used as they increase the execution speed of analysis of data applications tremendously. They are also proved to be highly scalable using additional nodes of clusters when the data volume is continuously increasing. These are also reliable as every bit of data is made a copy onto a different node ensuring no data loss on failure of one. This converges to the basic idea that data which is divided into parts can be exploited and used for bettering clustering process [2].

To make clustering more efficient, IFCM is better than FCM and has the intuitionistic feature in it, is used with Hadoop and it proves to be more promising for varied practical and business applications.

2 Existing Methods

Large amount of work has been done by researchers in the direction of clustering in big data. These algorithms in literature can be broadly divided and classified majorly into three categories. (a) CLARA [3], CURE [4], and the corset algorithms [5] are the sampling methods where cluster centres are size. (b) Incremental clustering [6], divide and conquer [7–9] represent single-pass algorithms in which data is loaded in sequence into data of small groups. These algorithms handle these parts and then combine the results. (c) Birch [10], Clarans [11], Garden [12] and Cluto [13] were successive in designing algorithms for handling high-dimensional data by transforming data so as to make it more accessible. The transformed data is usually represented by graphs. All the mentioned algorithms only produce crisp partitions.

Zadeh [14] brought in the concept of fuzzy sets. Fast FCM (FFCM) [15], and multistage random FCM [16] are good with respect to speed. However, they are not scalable. Some of the other algorithms mentioned in [17, 18] also add to the list of algorithms improving efficiency of the ones mentioned earlier. Fast Kernel FCM

[19] was developed for MRI image processing. Intuitionistic fuzzy sets were introduced by Attanassov [20]. Intuitionistic fuzzy C-means was introduced by Chaira [21, 22]. The supremacy of IFCM has already been established over FCM.

2.1 Fuzzy C-Means Algorithm (FCM)

FCM is an algorithm that was introduced by Bezdek [23]. The data elements in fuzzy or soft clustering can be members of more than one cluster. In addition, a set of membership level corresponding to each element is specified. This membership value helps in finding the extent of relationship between the element and that cluster.

1. Allocate initial means for c clusters.
2. Utilising Euclidean formula,

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Compute distance d_{ik} between x_k (data point) and v_i (cluster centroid).

3. Generate the fuzzy partition matrix or membership matrix U :

If $d_{ij} > 0$ then

$$\mu_{ik} = \frac{1}{\sum_{j=1}^C \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}} \quad (2)$$

Else

$$\mu_{ik} = 1$$

4. Compute v_i (cluster centroid) using

$$V_i = \frac{\sum_{j=1}^N (\mu_{ij})^m x_j}{\sum_{j=1}^N (\mu_{ij})^m} \quad (3)$$

5. Using Steps 2 and 3 compute new membership matrix.
6. If $\|U^{(r)} - U^{(r+1)}\| < \varepsilon$ then terminate otherwise repeat from Step 4.

2.2 Intuitionistic Fuzzy C-Means Algorithm (IFCM)

The IFCM introduced by Chaira [21] presents a new parameter denoted by π . This factor is known as hesitation value and it aids in increasing clustering accuracy.

1. Allocate initial means for c clusters.
2. Utilising Euclidean formula,

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4)$$

Compute distance d_{ik} between x_k (data point) and v_i (cluster centroid).

3. Generate the fuzzy partition matrix or membership matrix U :
If $d_{ij} > 0$ then

$$\mu_{ik} = \frac{1}{\sum_{j=1}^C \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}} \quad (5)$$

Else

$$\mu_{ik} = 1$$

4. The hesitation matrix π is calculated using

$$\pi_A(x) = 1 - \mu_A(x) - \frac{1 - \mu_A(x)}{1 + \lambda \mu_A(x)} |x \in X \quad (6)$$

5. Modified partition matrix U' is calculated using

$$\mu'_{ik} = \mu_{ik} + \pi_{ik} \quad (7)$$

6. Compute v_i (cluster centroid) using

$$V_i = \frac{\sum_{j=1}^N (\mu'_{ij})^m x_j}{\sum_{j=1}^N (\mu'_{ij})^m} \quad (8)$$

7. Using Steps 2–5 compute new membership matrix.
8. If $\|U^{(r)} - U^{(r+1)}\| < \varepsilon$ then terminate otherwise repeat from Step 4.

3 Proposed Method

In this paper, a novel algorithm is proposed in which data chunks are initially identified by Mapper class in Hadoop framework. The clusters along with arbitrary centroids serve as input to mapper class. The centroids are updated after each iteration of IFCM. This is done just after the reducer class merges the chunks. Applying IFCM alone on large data is less efficient and process becomes cumbersome. This drawback can be overcome using Hadoop. It helps tremendously for computing better clusters effectively and reach local optima efficiently. The algorithm, Hadoop Intuitionistic Fuzzy C-Means (HIFCM) algorithm proceeds as follows.

1. Random allocation of centroids is done based on the data set.
2. The input dataset contains complete vectors with centroid information without classes and is fed to the mapper class.
3. Mapper class is used to read the file and organise the data in chunks using appropriate data structures.
4. After reading is accomplished, mapper produces the nearest centroids using Steps 1–5 in IFCM. These values are dumped to reducer.
5. After collecting all the data, reducer emits the updated centroids.
6. Convergence is depicted if the distance between old and new centroids is less than 0.1 and process terminated. Else repetition is performed from Step 2 utilising updated centroids.

4 Experimental Results

4.1 Metrics and Evaluation Indices

The results of the proposed algorithm are evaluated using the metrics like accuracy, sum squares within (SSW), sum squares Between (SSB), Davies–Bouldin (DB) index, Dunn–Dunn index (DDI), and Silhouette coefficient (SC). Accuracy being one of the most important metrics helps in finding how correctly the clusters were formed.

SSW shows in what proximity is the sample to the cluster centre. SSB indicates the distance of the sample from the other cluster. DB which uses cluster centroids explains the dispersion within clusters and that with the other clusters. Smaller the DB value, better the clustering. DB measures how compact and separated the

clusters are from each other. DDI whose range lies in $[0, \infty]$, tells about separation ratio by increasing the distance between clusters and decreasing the distance within the clusters. SC evaluates the quality of each cluster by analysing the suitability of every object to its corresponding cluster. The values range from -1 to 1 with 1 indicating the best clustering [24].

The DBI and DDI form the primitive indices of performance analysis.

4.1.1 Davis–Bouldin (DB) Index

The DB index is defined as the ratio of sum of within cluster distance to between-cluster distance. It is formulated as given.

$$DB = \frac{1}{c} \sum_{i=1}^c \max_{k \neq i} \left\{ \frac{S(v_i) + S(v_k)}{d(v_i, v_k)} \right\} \quad \text{for } 1 < i, k < c \quad (9)$$

This index delivers minimum value if within cluster distance is less and between clusters separation is more. Therefore, a good clustering procedure results in a very low DB value [25].

4.1.2 Dunn-Dunn (DD) Index

It is computed using

$$Dunn = \min_i \left\{ \min_{k \neq i} \left\{ \frac{d(v_i, v_k)}{\max_l S(v_l)} \right\} \right\} \quad \text{for } 1 < k, i, l < c \quad (10)$$

A larger value for the DD index proves clustering to be more efficient [24].

5 Results

The proposed algorithm has been implemented in Java. In implementation, we have utilised **org.apache.hadoop.mapreduce** package. The experimentation was conducted on HP Pavilion g6-1a69us. Processor: Intel Core i5 380 M, 2.67 GHz, 2 MB cache memory, 4 GB RAM. It was having 64-bit Windows 7 Home and Eclipse Luna IDE.

Description of UCI Data Sets used

- Dataset 1: Credit data set—To assess application of credit cards based on the Australian credit card. It comprises 690 instances and 15 attributes.
- Dataset 2: Wine data set—Created as an outcome of chemical investigation of wines developed in a particular region. Total the dataset comprises of 178 instances and 13 attributes.
- Dataset 3: Glass data set—Attributes represent the type of oxide content. Total dataset comprises of 214 instances and 10 attributes.
- Dataset 4: Wisconsin diagnostic breast cancer (WDBC) data set—Comprises attributes that were extracted from a digitised image of a cell nucleus. It contained in total 569 instances and 32 attributes.

The clusters formed for wine and glass datasets with 3, 4, 5, 6 are as shown in Figs. 1 and 2.

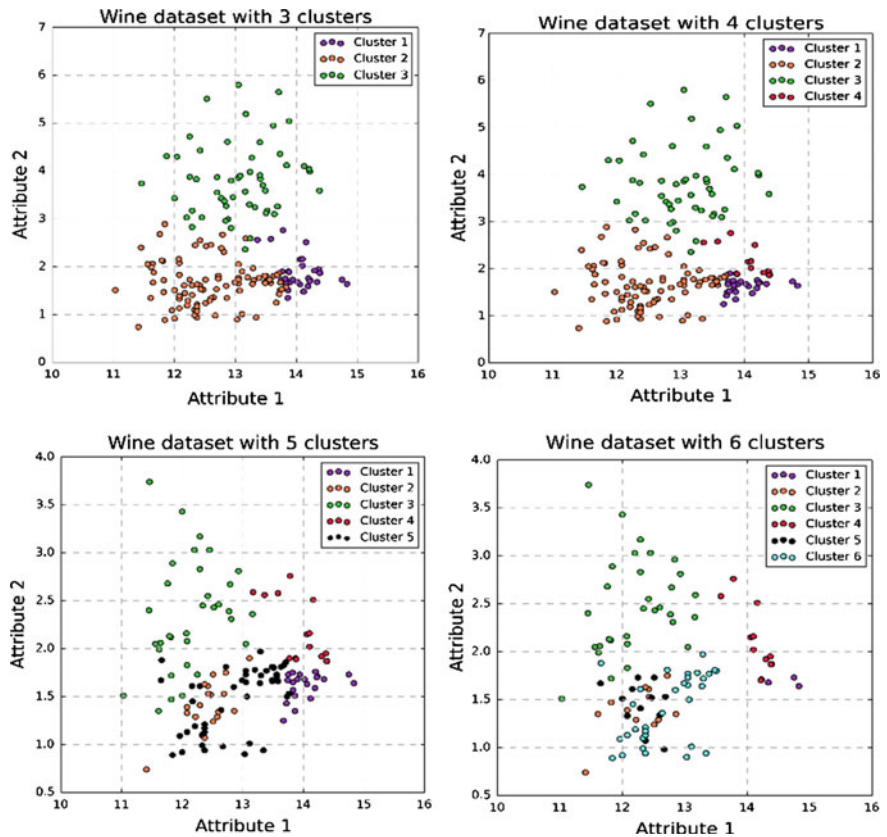


Fig. 1 Wine dataset with 3, 4, 5, 6 clusters

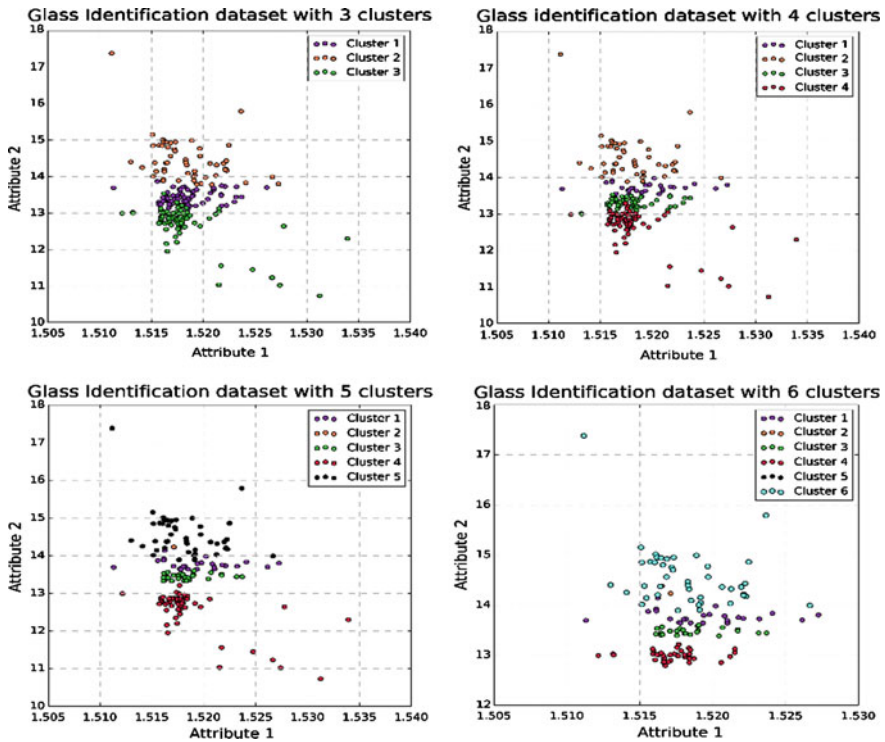


Fig. 2 Glass dataset with 3, 4, 5, 6 clusters

From the clustering in figures, we can infer that the clusters produced using HIFCM on large data are highly distinctive. The clustering is efficient with production of high-quality clusters. Clusters are dense and identifiable distinctly.

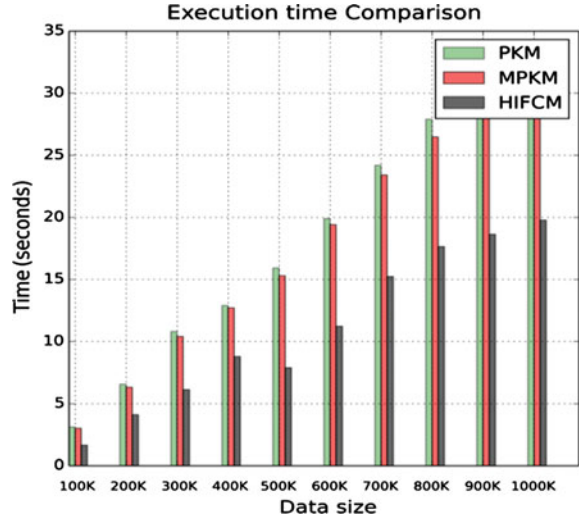
5.1 Execution Time

The execution times of HIFCM in comparison to PKM and MPKM algorithms proposed by Mathew et al. [26] are given in Table 1 and corresponding graph is

Table 1 Execution time of PKM, MPKM AND HIFCM

Data set size N		100 K	200 K	300 K	400 K	500 K	600 K	700 K	800 K	900 K	1000 K
Execution time	PKM	3.12	6.56	10.8	12.9	15.92	19.9	24.2	27.9	29.76	31.43
	MPKM	3.02	6.32	10.4	12.7	15.3	19.4	23.4	26.46	28.67	29.14
	HIFCM	1.65	4.11	6.12	8.78	7.89	11.23	15.23	17.64	18.63	19.76

Fig. 3 Execution time comparison of PKM, MPKM, HIFCM



shown in Fig. 3. The values have been computed by executing HIFCM keeping dimensionality of data, number of cores and data size uniform.

The values provided in Table 1 clearly imply that execution of HIFCM takes comparatively much less time than PKM and MPKM.

5.2 Performance Metrics

From the graphs, it is implied that the proposed HIFCM algorithm is more efficient in all aspects when compared to PKM and MPKM (Fig. 4).

The values in Table 2 clearly indicate that HIFCM is much better when compared to PKM and MPKM in terms of efficiency and accuracy. HIFCM has an advantage over PKM and MPKM as reflected by its decent values of SSB, DDI, and SC. Lower DBI of HIFCM indicates lesser dispersion. The larger value of DDI indicates higher intercluster distance and lower intracluster distance. The higher value of SC indicates good quality clustering compared to the other two algorithms. It can also be observed that SC value of HIFCM is closer to 1 indicating the effectiveness of HIFCM. In general scenario, for each dataset used, the HIFCM method outperforms PKM and MPKM techniques with respect to each of the validity measures.

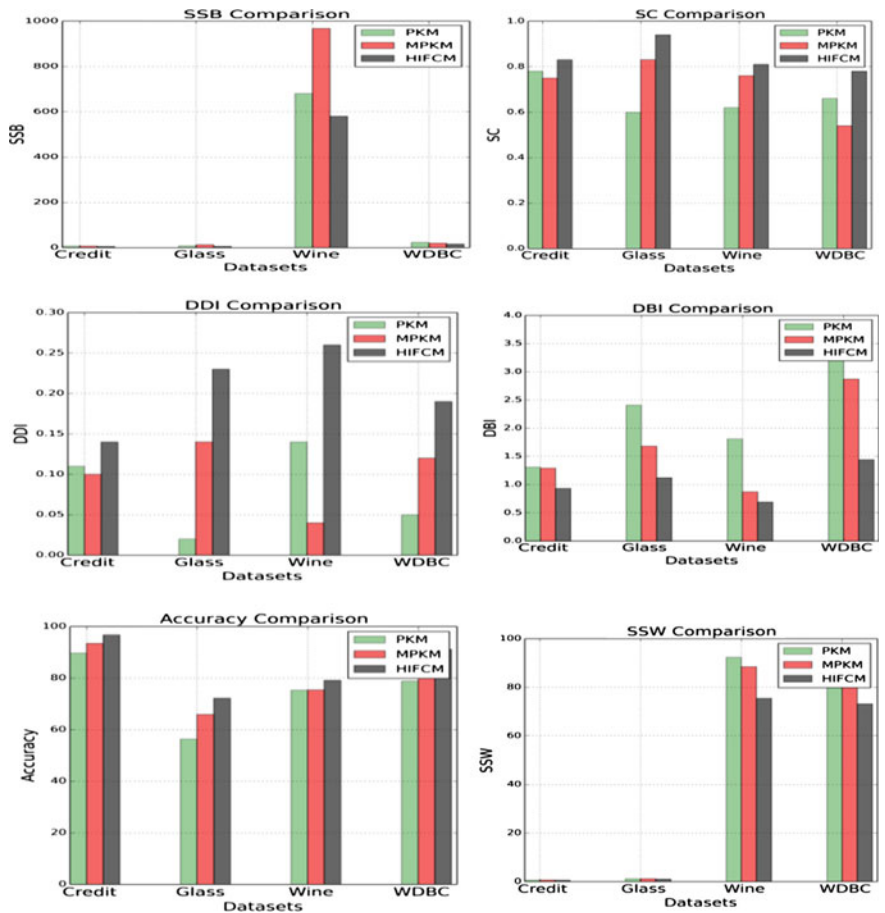


Fig. 4 Comparison of performance metrics of HIFCM with PKM and MPKM

Table 2 Cluster validity measures

Dataset	Algorithm	Accuracy (%)	SSW	SSB	DBI	DDI	SC
Credit	PKM	89.68	0.6578	6.89	1.31	0.11	0.78
	MPKM	93.45	0.6478	6.82	1.29	0.1	0.75
	HIFCM	96.32	0.61	6.13	0.84	0.14	0.838
Glass	PKM	56.32	1.1709	7.65	2.41	0.02	0.6
	MPKM	65.89	1.1821	12.69	1.68	0.14	0.83
	HIFCM	72.32	1.12	5.22	1.14	0.236	0.956
Wine	PKM	75.34	92.34	680.2	1.81	0.14	0.62
	MPKM	75.40	88.45	967.3	0.87	0.04	0.76
	HIFCM	79.39	78.34	540.628	0.61	0.269	0.827
WDBC	PKM	78.84	84.22	23.2	3.51	0.05	0.66
	MPKM	78.84	81.56	19.3	2.87	0.12	0.54
	HIFCM	91.46	74.43	12.427	1.42	0.19	0.783

6 Conclusion

The proposed Hadoop integrated IFCM (HIFCM) produces better results when compared to algorithms like parallel K-means and modified parallel K-means. This algorithm can be one of the solutions for solving various big data problems in industry and academia. However, various enhancements of the proposed algorithm can be done. The Euclidean formula is used to find the distance between points. This can be replaced with several other measures and hence can be tried to improve the effectiveness of the proposed algorithm.

References

1. <http://searchbusinessanalytics.techtarget.com/definition/Hadoop-cluster>.
2. <http://www.sas.com/resources/asset/five-big-data-challenges-article.pdf>.
3. L. Kaufman and P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley-Blackwell, 2005.
4. S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," *Inf. Syst.*, vol. 26, no. 1, pp. 35–58, 2001.
5. S. Har-Peled and S. Mazumdar, "On coresets for K-means and k-median clustering," in *Proc. ACM Symp. Theory Compute.*, 2004, pp. 291–300.
6. F. Can, "Incremental clustering for dynamic information processing," *ACM Trans. Inf. Syst.*, vol. 11, no. 2, pp. 143–164, 1993.
7. F. Can, E. Fox, C. Snively, and R. France, "Incremental clustering for very large document databases: Initial MARIAN experience," *Inf. Sci.*, vol. 84, no. 1–2, pp. 101–114, 1995.
8. C. Aggarwal, J. Han, J. Wang, and P. Yu, "A framework for clustering evolving data streams," in *Proc. Int. Conf. Very Large Databases*, 2003, pp. 81–92.
9. S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O'Callaghan, "Clustering data streams: Theory and practice," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 3, pp. 515–528, May/June 2003.
10. T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, 1996, pp. 103–114.
11. R. Ng and J. Han, "CLARANS: A method for clustering objects for spatial data mining," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 5, pp. 1003–1016, Sep./Oct. 2002.
12. R. Orlandia, Y. Lai, and W. Lee, "Clustering high-dimensional data using an efficient and effective data space reduction," in *Proc. ACM Conf. Inf. Knowl. Manag.*, 2005, pp. 201–208.
13. G. Karypis, "CLUTO: A clustering toolkit," Dept. Computer. Sci., Univ. Minnesota, Minneapolis, MN, Tech. Rep. 02–017, 2003.
14. L. A. Zadeh (1965) "Fuzzy sets". *Information and Control* 8(3) 338–353.
15. B. U. Shankar and N. Pal, "FFCM: An effective approach for large data sets," in *Proc. Int. Conf. Fuzzy Logic, Neural Nets, Soft Computing.*, Fukuoka, Japan, 1994, p. 332.
16. T. Cheng, D. Goldgof, and L. Hall, "Fast clustering with application to fuzzy rule generation," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Tokyo, Japan, 1995, pp. 2289–2295.
17. J. Kolen and T. Hutcheson, "Reducing the time complexity of the fuzzy C-means algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 263–267, Apr. 2002.
18. R. Cannon, J. Dave, and J. Bezdek, "Efficient implementation of the fuzzy C-means algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 2, pp. 248–255, Mar. 1986.

19. L. Liao and T. Lin, "A fast constrained fuzzy kernel clustering algorithm for MRI brain image segmentation," in Proc. Int. Conf. Wavelet Anal. Pattern Recognition., Beijing, China, 2007, pp. 82–87.
20. Atanassov, K.T.: Intuitionistic Fuzzy Sets, *Fuzzy sets and Systems* 20.1 (1986): 87–96.
21. Chaira, T. and Anand, S.: A Novel Intuitionistic Fuzzy Approach For Tumor/Hemorrhage Detection in Medical Images. *Journal of Scientific and Industrial Research*, 70(6), (2011).
22. Tripathy, B.K., Rohan Bhargava, Anurag Tripathy, Rajkamal Dhull, Ekta Verma, P. Swarnalatha: Rough Intuitionistic Fuzzy C-Means Algorithm and a Comparative Analysis in proceedings of ACM Compute-2013, VIT University, 21–22 August, 2013.
23. Bezdek, J.C., Ehrlich, R., and Full, W. (1984). "FCM: Fuzzy C-Means Algorithm" *Computers and Geoscience*, 10(2–3), 191–203.
24. Bezdek, J.C. and Pal, N.R.: "Some new indexes for cluster validity", *IEEE Transaction on System, Man and Cybernetics, Part B: Cybernetics*, vol. 28, pp. 301–315, 1998.
25. Davis, D. L. and Bouldin, D.W.: "A cluster separation measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, (1979), pp. 224–227.
26. Mathew J and Vijaykumar R, "Scalable parallel clustering approach for large data using parallel K means and firefly algorithms", *IEEE, Proceedings of International Conference on High Computing and Applications*, 2014, pp. 1–8.

A Framework for Group Decision Support System Using Cloud Database for Broadcasting Earthquake Occurrences

S. Gowri, S. Vigneshwari, R. Sathiyavathi and T.R. Kalai Lakshmi

Abstract The proposal deals with designing of a novel framework for group decision support system (GDSS). The purpose of applying such a system is to broadcast the events of an earthquake occurrence based on group decisions by analyzing the reports from the sensor networks which are set in the deep sea. Recent advancements in the field of data analytics show that automatic decision making system is indispensable, exclusively in case of an emergency circumstance. Current frameworks are mostly based on single-user decision support systems. Well-established operational research techniques such as multi-criteria decision making techniques need to be integrated in order to bring out a successful decision support system. A comprehensive analysis of the challenges in the fields of information systems and machine learning frameworks is to be scrutinized. Such a study is critical for designing a framework based on platform independent automated multi-decision-based GDSS. This can efficiently support alternative types of goals and control protocols between its users.

Keywords Sensor network · Group decision support system · Decision making framework · Operational research techniques

1 Introduction

A computer-based information framework in operational research techniques that supports decision making activities is called the decision support system (DSS). DSSs provides the administration, functionalities, and development levels of an organization (in our GDSS framework government is responsible) and assist public make decisions about harms that may be quickly varying and not effortlessly specified in prior [1, 2]. Types of decision support systems are as listed below:

S. Gowri (✉) · S. Vigneshwari · R. Sathiyavathi
Faculty of Computing, Sathyabama University, Chennai 600019, India
e-mail: gowriamritha2003@gmail.com

T.R. Kalai Lakshmi
Faculty of Business Administration, Sathyabama University, Chennai 600019, India

- Fully human-powered
- Computerized [3]
- Combination of both human-powered and computerization.

Academics perceive DSS as a tool to support decision making procedure; DSS is seen as a tool to facilitate organizational processes by the DSS consumers.

DSS being a traditional approach is in consideration with the current scenarios, a novel approach of group decision support system (GDSS) is proposed in this paper. GDSS is the decision making approach utilizing multiple decisions from multiple inputs (humans, computers, or both) [4, 5]. GDSS results in much better way of resultants rather than singly decided approaches.

2 Proposed Work Description

The proposed work is an integrated frame of the individual units below:

1. Network of sensors in sea
2. Centralized cloud server database and knowledge
3. Broadcasting tower
4. Central government hubs
5. Information transmission unit.

Figure 1 depicts an overview of the proposed framework. The sensor network, deep inside the sea will sense the underwater earthquakes and it transmits the signal to the centralized cloud server. The cloud server in turn will broadcast the signals to the decision group. The decision group comprises of government bureaus, social

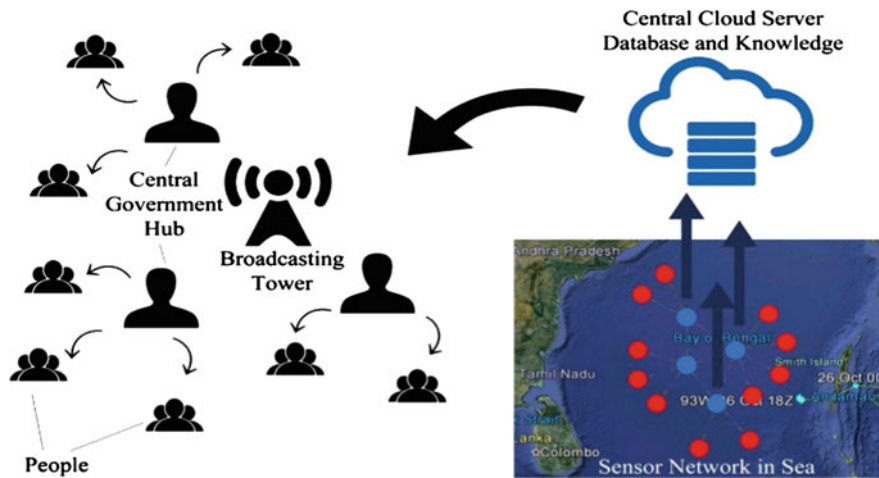


Fig. 1 An overview of the proposed framework

networking groups, emergency utilities, etc. Based on the responses from the user groups, the system will generate automated decisions in order to carry out necessary actions and announcements.

Description of each unit in detail is given below.

2.1 Network of Sensors in Sea

Wireless sensor networks (WSN) are spatially dispersed self-ruling sensors to screen physical or natural conditions, for example, temperature, sound, weight, and so on and to helpfully pass their information through the system to a primary area. The more advanced systems are bidirectional, furthermore empowering control of sensor action.

The WSN is constructed of “hubs”—from a couple to a few hundreds or even thousands, where every hub is associated with one (or sometimes a few) sensors. Each such sensor system hub has regularly a few sections: a radio handset with an interior receiving wire or association with an outside reception apparatus, a microcontroller, an electronic circuit for interfacing with the sensors and a vitality source, for the most part a battery or an inserted type of vitality reaping. The topology of the WSNs can fluctuate from a basic star system to a progressed multi-bounce remote cross section system. The spread system between the bounces of the system can be steering or flooding.

2.2 Centralized Cloud Server Database and Knowledge

A cloud database is a database that regularly keeps running on a cloud computing stage. There are two basic organization models: clients can run databases on the cloud autonomously, utilizing a virtual machine picture, or they can buy access to a database management, reserved up by a cloud database supplier.

Cloud storage is a model of information supplying where the advanced information is put away in sensible pools, the physical supplying compasses different servers (and frequently areas), and the physical environment is ordinarily possessed and oversaw by a facilitating organization. These cloud storage suppliers are in charge of keeping the information reachable and obtainable, and the physical environment secured and running. Individuals and associations purchase or lease supply limit from the suppliers to store client, association, or application information.

Cloud storage administrations may be gotten to through a cofound cloud PC administration, a web administration application programming interface (API) or by applications that use the API, for example, cloud desktop supplying, a cloud storage passage, or web-based substance administration frameworks.

2.3 *Broadcasting Tower*

A common broadcast tower on land would broadcast information present in the cloud database to the Internet hubs in the government in order to transmit information from the network sensors. According to which the next process would be carried on.

2.4 *Central Government Hubs*

The central government hubs are present in the government offices where the information which is been collected and noted, based on this collected information from various sensor networks present in the sea, the decision is made automatically as well as manually and then under the government authentication the information is made to be transmitted to the public through information transmission unit [6].

2.5 *Information Transmission Unit*

The information transmission unit has the operation of information passage to the public through various medium of transmission, such as mobile phone, televisions, and so on [7].

A simulation of the network model was done utilizing MATLAB 2009b by manually setting the parameter of the nodes. The simulated graph is shown in the Fig. 2.

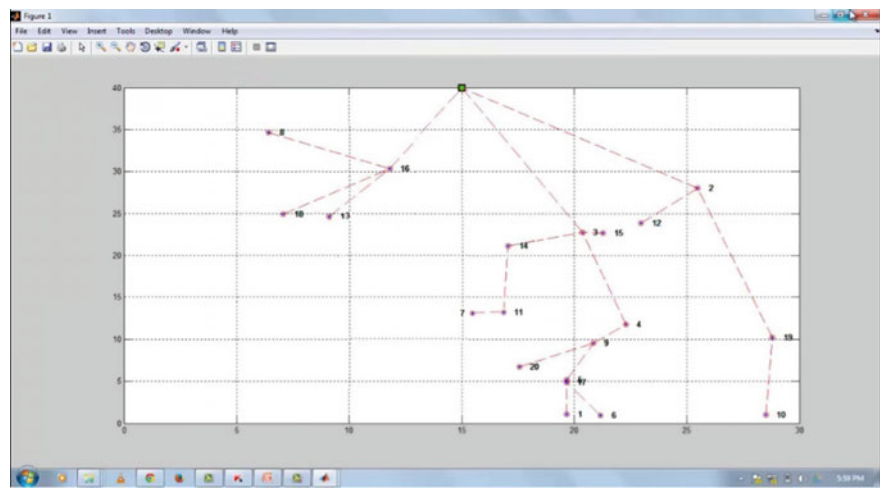


Fig. 2 Sensor network simulation

The decision making analysis was done but in computerized manner as well as manually. The computerized decision making was done using the organization of data according to the relevancy of the data collected from sensor network and then future display in categories accordingly.

The casting was done using a mobile which acted as a broadcaster where an application was designed to pass messages according to the finalized decision. The application has a database connected in which all the contact numbers are availed in the database.

3 Conclusion

The proposed system had resulted in an immediate message transformation process. The information was sent from the simulated network to the cloud database and after which the data was transmitted to the government hub using the broadcasting tower. The transmitted data is then sent to the decision making unit where the data is categorized according to the relevancy of the data. The categorized data is then manually monitored and passed to the public. The set system was scrutinized to the extreme on its working, further evaluation must be done.

References

1. Turban, E., Aronson, J. & Liang, T. (2005). "Decision Support Systems and Intelligent Systems", Upper Saddle River: Pearson Prentice Hall.
2. F. G. Filip, (2008) "Decision support and control for large-scale complex systems", *Annual Reviews in Control*, vol. 32, no. 1, pp. 61–70.
3. L. Duta, A. Bituleanu, F. G. Filip and I. Istudor, (2009) "Computer-based decision support for railroad transportation systems: an investment case study", *Informatica Economica*, vol. 13, no. 2, pp. 103–109.
4. Patrick, H. & Garrick, J. (2006, April). "The Evolution of Group Decision Support Systems to Enable Collaborative Authoring of Outcomes" *World Futures: The Journal of General Evolution*. 62(3), 193–222.
5. Boutari, C. (2004) "'Opening Up' Group Decision Support (GDS): An exploratory Study of Contemporary to GDS", London Multimedia Lab RR 04/02, London School of Economics and Political Science.
6. Jones, G (2005) "Learning environments for collaborative authored outcomes", Research Report LML-LUDIC-RR-002. London Multimedia Lab for Audiovisual Composition and Communication.
7. Khourian, S., (2004) "Creative Mediation", in *With All Due Intent*, Catalogue of Manifesta 5, San Sebastian.

Advanced Persistent Threat Model for Testing Industrial Control System Security Mechanisms

Mercy Bere-Chitauro, Hippolyte Muyingi, Attlee Gamundani
and Shadreck Chitauro

Abstract An APT is a targeted multi-step attack that uses zero day exploits to achieve its objectives. In order to find solutions to mitigate APT attacks it is important to understand APT anatomy. This paper proposes an APT testing model developed using design research methodology that can be used to develop industrial control security (ICS) mechanisms. The model development followed three steps; identifying the components; identifying and explaining the characteristics in each component and developing the model. Six components were identified to be included in the model; reconnaissance, injection, installation, operation, command and control and termination. The model proposed is envisaged as systematic approach to testing and validation of security mechanisms that are aimed at APT detection in ICS.

Keywords Advanced persistent threats · Industrial control system · Security · Attacks · Threats

1 Introduction

Advanced persistent threats (APT) are persistent cyber-attacks that stealthily infiltrate a network [1]. APT use reconnaissance attacks to gain information about their targeted networks. The information from the reconnaissance attack is used to find

Mercy Bere-Chitauro (✉) · Hippolyte Muyingi · Attlee Gamundani · Shadreck Chitauro
Computer Science Department, Namibia University of Science and Technology
(formerly Polytechnic of Namibia), Windhoek, Namibia
e-mail: mbere@polytechnic.edu.na

Hippolyte Muyingi
e-mail: hmuyingi@polytechnic.edu.na

Attlee Gamundani
e-mail: agamundani@polytechnic.edu.na

Shadreck Chitauro
e-mail: schitauro@polytechnic.edu.na

ways and methods to gain access into the system. Once an APT has found its entry point and positioned itself strategically in the network it establishes a communication channel with its command and control centre. Updates on the APT and further instructions are sent from the command and control centre. Information gathered about the system by the APT is also sent to the command and control centre [2, 3]. APTs are hard to detect in a network because they use sophisticated techniques to camouflage their activities from the usual detection mechanism systems in networks. APTs can also attribute their success to zero-day exploits that are usually entrenched in them to attack networks for the sole purpose of not being detected by intrusion detection systems and antiviruses which normally rely on previously known signatures [4].

The APT; Stuxnet which was discovered in 2010 used zero-day exploits to sabotage Iranian Natanz Nuclear Enrichment facility operations an industrial control system facility [1]. Miniduke a more recently discovered APT which was first announced by FireEye on February 13, 2013 was also targeting industrial control system facilities [5].

ICS are used to automate distribution of water, electricity, natural gas, oil. They are used to regulate and manage industrial processes like food, beverage and pharmaceutical manufacturing. Furthermore they control air traffic, materials handling, postal mail handling, railway transportation systems, communication networks and mining industries, wastewater treatment and specialised facilities like nuclear plants. Successful APT attacks in ICS will result in endangering people's health and safety [6] and successful attacks might result in damage to infrastructure and in most instances there will be a financial losses as a result [6].

As the number of APT being discovered is increasing [7] it is vital to design security mechanisms that ICS can use to detect and to stop APT attacks. The overall research objective is to design a bio-immunology inspired ICS security model to improve ICS defence from APT but as was outlined APT are always unique and target specific organisations and this makes it impossible to have a one size which fits all APT to use to test for APT detection and deterring security mechanism. It is a common knowledge that an APT is executed in multiple stages and as such it is important to establish common characteristics and traits at each stage of an APT attack so that whenever an APT security mechanism is being tested the testing tools should have those characteristics. By incorporating the common characteristics it will be possible to sufficiently test that particular security mechanism at each stage of the APT attack to deduce whether the security mechanism is effective or not. It would not make sense to use known APT for testing the effectiveness of the security mechanisms because APT are designed differently for a particular organisation so an APT used to attack a certain organisation may never be reused again.

The paper is organised as follows; the next section will outline the methods used to design APT model. Section 3 will describe the actual APT development process and the components identified. Section 4 will explain the APT model and its components. Section 5 will discuss how the APT model will be used for testing and finally Sect. 6 will be the conclusion and some remarks on future research.

2 Methodology

The overall objective of the project is to design the bio-immunology-inspired networking security model for ICS defence from APT. To design the model design research methodology is being used because it emphasises development of new artefacts and outcomes that solve current problems [8]. Hevner et al. [9] state that it is useful in solving ‘wicked problems’ characterised by among others:

- A critical dependence upon human cognitive abilities (e.g., creativity) to produce effective solutions, and
- A critical dependence upon human social abilities (e.g., teamwork) to produce effective solutions.

Design science guidelines proposed by [9] were considered in particular that “Design science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation”. The three proposed by [10] shows that there are three design science research cycles: relevance cycles, design cycle and rigor cycle. In the relevance cycle one needs to identify the research requirements. In this instance the requirements was to have APT to test usefulness of bio-immunology-inspired networking security model. To do this, there is a need to have a way to test with a complete APT which we perceived would be realistic if we model the APT in stages.

In the rigor cycle, knowledge is brought about from existing knowledge which was done through the use of qualitative research. In this case, a literature review of APT was done to come up with themes, patterns or categories associated with APT. Design cycle cements all cycles in that requirements for the design cycle emanate from relevance cycle and evaluation and design theories are taken from the rigor cycle.

3 Designing APT Testing Model

The model development had three steps; identifying the components through literature study; identifying and explaining the characteristics in each component through analysis of APT behaviour through literature, and developing the model.

Current research [2–4] describes APT as sophisticated multistep cyber attacks. They state that an APT has these stages; choosing a victim, reconnaissance, delivery, exploitation, operation, data collection and exfiltration.

The first stage of an APT attack is deciding on a victim and mapping out the objectives of that attack [3]. In the reconnaissance stage, the attackers gain information about their target by doing network scanning and using social engineering techniques [2, 3]. In the next stage, the APT is delivered in the system mostly using social engineering especially spear phishing emails to the target organisation or they use infected USB stick [11]. The next stages of exploitation and operation involve exploiting vulnerabilities to execute APT payload and to stealthily and persistently

Table 1 APT permutations example

APT stage	Reconnaissance	Delivery	Operation	Command and control
Programs available	A, B, C	D, E, F	G, H, I	J, K, L

maintain presence in the system [2, 3, 11]. Finally or during APT operation data which is stolen is encrypted and sent to command and control centres. After being discovered most APT shut down their command and control centres to cover their tracks [12–14].

In testing for the different ways to deter and detect APT, most research use an APT samples that incorporates all the stages of an APT in one instance. This paper is proposing to design an APT testing model to effectively test for security mechanisms that will best detect APT in organisations where there are limited resources and where it is not possible to invent a new APT. The APT stages will be the components of the model that will be explained by their characteristics. When it comes to testing then the use of programs that will have those characteristics will be used for testing for each stage of the APT. Since APT are not designed in the same manner using a *program A* that mimics delivery of an APT and *program D* that mimics operation and then in the other instance using *program B and C* will make the APT capabilities unique in each scenario. To clarify consider the Table 1.

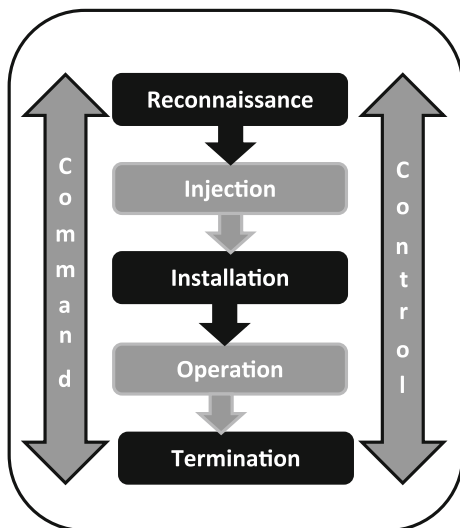
It is possible to have an APT attack consisting of programs A, D, G, J or A, F, G, L or C, E, G, K. All in all 495 permutations are possible.

Based on the analysis of the literature describing APT, this research identified the following APT elements which will be the components of the model. The components need to be related in order to make out a complete APT. The models’ components identified are

- Reconnaissance
- Injection,
- Installation,
- Operation,
- Command and control,
- Termination.

4 APT Testing Model

The model presented below is based on the literature review on APT attacks life cycles as explained by [15] where the main focus was to try and present a detailed analysis of the nature of APT attacks, their anatomy and in relation to how possibly they can be used to test ICS security mechanisms that should prevented them to a certain extent. Clearly the model presented in Fig. 1 is still at the theoretical level

Fig. 1 APT testing model

and subject to practical testing hence a proposal. As hinted by [11], a new approach is needed that takes a stepwise formality and be able to link analysis methods to attack features, which is the basis of this model design.

4.1 Reconnaissance

At this stage, the investigators will learn about their target as informed by the behaviour of attackers. The norm of making use of the weakest link to attack a given network as explained by [16], is what has to be modelled at this stage. This might involve the use of network scanning and mapping mechanisms to gather information or social engineering techniques, employee profiling, social networks and phone directories to get information. At this instance, the investigators will try to find ways of delivering APT into a system without being easily detected, as [17] clearly explains on the behaviour of attacks that perpetuate stealth operations [17].

To embrace the greater scope of the target, an approach as proposed by [18] that keeps track of system events, their dependencies and occurrences, is mandatory to formulate at this stage. The main reason for this extended approach is motivated by what [19] present in light of critical infrastructures becoming too interconnected and the utilisation of off-the-shelf application packages, that may make the reconnaissance stage much more complex to handle.

4.2 Injection

This stage will entail the use of the information gathered in the reconnaissance stage to exploit the identified vulnerabilities to deliver software that will be used in the next stage. The software should be able to communicate with command and control. The concept of reverse engineering as presented by [20] in the context of imitating the attack of botnets will prove handy to a certain level. The reconnaissance populated data will determine the injection techniques to be employed for modelling a particular APT attack.

4.3 Installation

This stage will involve the use of the information gathered in the reconnaissance stage to exploit the identified vulnerabilities to install software injected in the system in the previous stage. The software should be able to communicate with command and control. The need to communicate with the command and control wings is explained well by [21] where the dynamism and heterogeneity of APT is emphasised.

Some environments may demand different installation mechanisms unlike others that are routine as attacks peculiar to smart grids which have well-specified procedures, which conclusively presents them as easily predictable and routine [18].

4.4 Operation

At this stage software installed should stealthily maintain presence in the system, and or compromise the system and or gather data about the infected system and or escalate privileges. The software should be able to communicate with command and control. The operation stage should aim to portray a variety of components and techniques that work in synchrony to imitate the deployment stature of APT on an identified target [17].

The operation phase should handle both the pre-intrusion and post-intrusion as that determines the persistence nature of the APT attack. The built of a model such as one presented in this paper is guided by hints given by [22] on the need to take care of complex networked systems.

4.5 Termination

This stage should involve stopping the APT operations and stoppage procedures like wiping memory either by itself or through communication from command and control.

4.6 Command and Control

This will be the central point where APT attacks are launched from or were data stolen should be sent to. The anatomy of APT as postulated by [2] promotes the need to have a guided functional bisection of the various components that built them and make them thrive. The command and control acts as shields as the approach presented in Fig. 1 is a risk-based approach to risk [23]. If the testing environment is not contained, it can also become a source of uncontrollable threats. The command and control arms make it possible to model the sophisticated security threats of APT which has been swelling in scale for decades [19].

The fact that detecting and defending APT is not easy as some are multi-staged [11], necessitates the need to have the command and control shields for a well curated environment.

5 Application of the APT Model

The proposed APT model for testing security mechanisms being developed to detect and stop APT is a flow progression. The users who need to test their new ICS security mechanism against APT will find this manageable because testing the APT will be conducted in a progressive manner in their systems. By testing their security mechanisms from reconnaissance, injection, installation, operation to termination and checking to see if there are any positive results of detection or stoppage they will be in position to sufficiently test how secure the new system is.

We thus propose a flow progression of the model when using it. That is, do reconnaissance then, inject malicious software, install it, operate the software, terminate or persist after execution and in all stages communicate with command and control.

The APT testing model is developed to assist in the development of tools or mechanisms that can be used to deter and detect APT. Thus, if the developer of an ICS security tool can manage to have all those stages in their tests then they can yield holistic results about how an APT will be handled by their innovation. It is important to repeat the different stages with a combination of different software or procedures so that it can be inferred that a different APT was used for each iteration.

6 Conclusions and Future Work

This paper proposed the use of APT testing model to help in testing ICS security mechanism that are being developed to secure ICS from APT. Most APT are unique so in trying to test security solutions against APT it is hard to find new APT that can be used to test in instance where there are few resources. Thus, if several iterations of different parts of the model are used with different software then it can be inferred that the software combination that was used was not the same and thus will entail a different APT at the end of each cycle.

In the future, we foresee that this model can be used to provide a systematic approach to testing and validation of security solutions that are aimed at APT detection.

References

1. E. Knapp.: Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, Scada, and other Industrial Control Systems. Ebook. Elsevier, Waltham (2011).
2. P. Giura and W. Wang.: A Context-Based Detection Framework for Advanced Persistent Threats. In: 2012 International Conference on Cyber Security, pp. 69–74. IEEE Xplore Digital Library (2012).
3. Threats on the Horizon- Rise of Advanced Persistent Threats. <http://www.fortinet.com/sites/default/files/solutionbrief/threats-on-the-horizon-rise-of-advanced-persistent-threats.pdf>.
4. Virvilis, N., Gritzalis, D., Apostolopoulos, T.: Trusted Computing versus Advanced Persistent Threats: Can a Defender Win This Game?. In: 10th International Conference on Autonomic and Trusted Computing (UIC/ATC), pp. 396–403. IEEE Xplore Digital Library (2013).
5. Kaspersky Lab Identifies ‘MiniDuke’, a New Malicious Program Designed for Spying on Multiple Government Entities and Institutions Across the World. http://www.kaspersky.com/about/news/virus/2013/Kaspersky_Lab_Identifies_MiniDuke_a_New_Malicious_Program_Designed_for_Spying_on_Multiple_Government_Entities_and_Institutions_Across_the_World.
6. Hadziosmanovic, D., Bolzoni, D., Etalle, S & Hartel, P.: Challenges and Opportunities in Securing Industrial Control Systems. In. Complexity in Engineering (COMPENG), pp 1–6. Xplore digital library. (2012).
7. State of Endpoint Risk. <https://www.lumension.com/Lumension/media/graphics/Resources/2014-state-of-the-endpoint/2014-State-of-the-Endpoint-Whitepaper-Lumension.pdf>.
8. Hevner, A., Chatterjee, S.: Design Research in Information Systems. Springer, USA (2010).
9. Hevner, A., R.; March, S.T.; Park, J. Ram. S.: Design Science in Information Systems Research. <http://www.brian-fitzgerald.com/wp-content/uploads/2014/05/Hevner-et-al-2004-misq-des-sci.pdf>.
10. Hevner, A., R.: A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems. 19 (2), 1–6 (2007).
11. Bhatt, P. Toshiro Yano, E., Gustavsson, P.M.: Towards a Framework to Detect Multi-stage Advanced Persistent Threats Attacks. In. 8th International Symposium on Service Oriented System Engineering (SOSE), pp 390–395. IEEE Xplore Digital Library (2014).
12. Targeted cyberattacks Logbook. <https://apt.securelist.com/#firstPage>.
13. COSMICDUKE Cosmu with a twist of MiniDuke. https://www.f-secure.com/documents/996508/1030745/cosmicduke_whitepaper.pdf.
14. Shmoon, a two-stage targeted attack. <http://www.seculert.com/blog/2012/08/shmoon-two-stage-targeted-attack.html>.

15. Ask, M., Bondarenko, P., Rekdal, J. E., Nordbø, A., Bloemerus, P., & Piatkivskyi, D. (2013).: Advanced Persistent Threat (APT) Beyond the Hype. Project report in IMT4582 Network Security at Gjovik University College, Springer.
16. Advanced malware detection through attack life cycle analysis. http://www.isc8.com/assets/files/CyberadAPT.WhitePaper.7000_HN.pdf.
17. Sood, A. K., & Enbody, R. J.: Targeted Cyberattacks: A Superset of Advanced Persistent Threats. *IEEE Security & Privacy*. 1, 54–61 (2013).
18. Skopik, F., Friedberg, I., & Fiedler, R.: Dealing with Advanced Persistent Threats in Smart Grid ICT networks. In: *Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5. *IEEE Xplore Digital Library* (2014).
19. Juuso, A. M., Takanen, A., Kittilä, K.: Proactive Cyber Defense: Understanding and Testing for Advanced Persistent Threats (APTs). In: *12th European Conference on Information Warfare and Security (ECIW)*, pp 383. *Academic Conferences Limited*. Chapter 2. (2013).
20. Binsalleeh, H., Ormerod, T., Boukhtouta, A., Sinha, P., Youssef, A., Debbabi, M., & Wang, L. On the Analysis of the Zeus Botnet Crimeware Toolkit. In: *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on* (pp. 31–38). *IEEE Xplore Digital Library* (2010).
21. Okhravi, H., Haines, J. W., Ingols, K.: Achieving Cyber Survivability in a Contested Environment Using a Cyber Moving Target,” *High Frontier Journal*. 7(3), 9–13 (2011).
22. Kato, M., Matsunami, T., Kanaoka, A., Koide, H., & Okamoto, E.: Tracing Advanced Persistent Threats in Networked Systems. In *Automated Security Management* (pp. 179–187). *Springer International Publishing*, (2013).
23. Cole, E.: *Advanced persistent threat: understanding the danger and how to protect your organization*. Newnes, (2012).

e-Reader Deployment in Namibia: Fantasy or Reality?

Mohammed Shehu and Nobert Jere

Abstract As technology grows and the usage of ICTs become popular all over the world, modern approaches are required to improve service delivery through technology. The education sector has been a beneficiary of such approaches. In some cases, however, determining the best technologies to implement have proven a much more complex endeavor. In this paper, an evaluation of the possible impacts of e-reader implementation in Namibian schools was undertaken, with several stakeholders within the Namibian education sector engaged and their feedback recorded and analyzed. Findings show that several infrastructural and social issues still need to be addressed before the application of innovative solutions to existing challenges, such as e-reader deployment in schools, can succeed.

Keywords ICTs · e-readers · e-learning · Implementation strategy

1 Introduction

Several arguments exist for the integration of ICT into education. These include improved educational outcomes, empowerment of disadvantaged population segments, and both increased and equitable access to ICT services. Namibia has embraced the use of ICT's in various sectors of the economy. Such ICT developments have been implemented within the education sector to improve teaching and learning.

Mohammed Shehu (✉) · Nobert Jere
Polytechnic of Namibia, Windhoek, Namibia
e-mail: mib.shehu@gmail.com

Nobert Jere
e-mail: njere@polytechnic.edu.na

Despite significant investment into Namibia's education sector, however, the sector still experiences poor infrastructure and high technological illiteracy [1]. Underutilization of ICTs, high technological illiteracy, and poor ICT implementation strategies also persist.

In this paper, we evaluate the current state of education in Namibia and assess the capability of e-readers to help in combating certain challenges. The benefits of e-readers include portability, low energy consumption, increased capacity for educational content storage at no extra weight, low pricing and Wi-Fi connectivity. E-readers could be utilized in Namibia's education sector to solve some current problems, especially the lack of teaching and learning materials. We also outline appropriate implementation strategies for successful e-reader integration into Namibian schools.

2 Overview of ICTs in Namibia

2.1 NIED Policy

The National Institute for Educational Development (NIED) consults on and publishes the national school curriculum for different levels including primary and secondary. The NIED acknowledges that ICT integration carries a complexity that necessitates clear-cut guidance [2]. The policy defines several goals, including producing tech-literate graduates, making educational institutions more efficient and effective, as well as broadening access to quality education at all levels. Angula and Mutorwa [2] recognize several fundamental factors that will play a significant role in the transformation:

- **Staff Training**—The NIED recognizes that preparing teachers to teach using ICT is a necessary step [2].
- **ICT Services**—These services include development, distribution, maintenance, and support of ICT.
- **Curriculum**—A change in the curriculum for basic education is cited as another necessary cornerstone of transformation.
- **Performance Measures**—The NIED Policy on ICT further recognizes the need for constant assessment.

Several challenges exist facing the implementation of ICT's in Namibia. These are:

- **Level of Research Maturity**—The very small number of research institutions and staff exacerbates the research capacity shortfall in the country [3].
- **(Un)Sustainable Spending**—The Namibian government devotes a significant percentage of its budget every year into various education-based initiatives; [4, 5].

- **Infrastructure**—Vast geographical distances contribute to the lack of ICT penetration and adequate electricity grid coverage in rural areas [6].
- **Training**—Implementation of ICT solutions requires extensive training of key stakeholders in relevant sectors.

3 E-Reader Research Techniques

Being a body of research primarily concerned with the introduction of new technological tools into the education sphere, it is most appropriate that it should be conducted under the umbrella of design-based research, while analysis is carried under grounded theory.

3.1 *Defining Design-Based Research*

Design-based research focuses on solving broad-based, complex, and real-world problems that are critical to education, with the end goal of making contributions both scientific and applied to the field [7–9]. Van den Akker [8] points out that in DBR output is measured as design principles, aiming to benefit all stakeholders involved [10].

3.2 *Phases of Design-Based Research*

- PHASE 1: **Problem definition**—In this phase, researchers and educators collaborate to identify and define practical problems within education, including a review of relevant literature [8, 9, 11, 12].
- PHASE 2: **Theoretical framework definition**—This involves defining the theoretical framework (preferably pragmatic-based) that will underpin the research to be undertaken [9, 13, 14].
- PHASE 3: **Iterative testing**—This involves the selection of methodologies (either quantitative or qualitative), the participants, and iterations of intervention implementation [9, 15].
- PHASE 4: **Production of design principles**—This sees the distillation and analysis of received data into key design principles to be used to guide future implementations and policymaking [9].

4 Methodology

Questionnaires were handed out to learners and teachers in three different schools within the Windhoek City area, as well as interviews with school administrators and management staff at one of the leading Namibian publishing houses.

4.1 Data Analysis

Data analysis was carried out using grounded theory, allowing themes, issues, and important topics to emerge from the data through iterative analysis of said data; these topics then form the basis for subsequent analysis [16].

5 Findings and Aggregated Analysis

Learners in Namibian classroom exhibit poor reading skills generally in and out of the classroom. They get distracted easily by their cell phones and social networking apps, and are not granted sufficient reading time during the week. They exhibit low attention spans when reading and feel that they would concentrate more if the content was more interesting. They prefer reading physical textbooks, but would not mind migrating all their textbooks to a digital device.

Teachers in Namibian agree that a digital solution to current problems would go a long way to help teaching, but that such a solution would require extensive teacher training.

Implementation is also hindered by other obstacles, such as poor basic infrastructure in underserved communities and the social backgrounds of many learners that exacerbates a poor reading culture. Improving the life of learners and their families on a macro-level would provide the necessary financial, technological, and intellectual support structure to justify the implementation of such future-forward technologies. A specific focus on community libraries and reading mentorship programs would ensure that early age learners got the necessary skills required to excel in school and integrate into society later in life. Funding for such initiatives would need to be sourced from a combination of government and public-private partnerships.

The following chapter presents a possible e-reader implementation strategy, outlining the relevant roles and stakeholders.

6 Implementation Strategy and Framework

Defining the roles that relevant stakeholders will play enables the creation of a robust strategy for e-reader deployment. The role-definition process was supported by careful consideration of existing literature pertaining to ICT implementation in schools, the identification of pivotal stakeholders in the Namibian education sector, the engagement of these stakeholders through qualitative methods and a comparison of findings from the research to the existing literature.

6.1 *Defining Stakeholder Roles*

- **Government**—Government would be responsible for building and maintaining the requisite infrastructure such as electricity coverage and internet access in all schools, as well as providing teacher training and subsidies for device acquisition.
- **Teachers**—Teachers would be tasked with developing course syllabi, undergoing professional training and liaising with content publishers.
- **Publishers**—Publishers would design and develop digital content and layout for the devices, as well as publishing and marketing quality content to give teachers a choice in teaching material. Cost-effective publishing practices would be a priority in order to keep final costs down.
- **Community leaders and administrators**—These stakeholders would provide awareness of new teaching methods, support, and engagement to schools during the transition, as well as encouraging and rewarding content developers to come up with quality content for the devices, thus building local capacity.
- **Software developers**—Developers would design, develop and publish engaging, interactive and educational apps that can be integrated into the learning process.
- **Learners**—Finally, learners would simply be required to use the devices and all apps, tools, and resources within the learning ecosystem to improve all necessary metrics such as reading literacy, writing ability, numeracy skills, and subject comprehension.

These roles can play a central role within a larger implementation framework. It is evident from the findings that different factions have different views on the viability and strategy of e-reader implementation in the country. The following encompassing framework is derived from the amalgamation of findings from this research, as well as from principles culled from the literature. It is implied that successful e-reader deployment will have to consider the key factors such as policy requirements, economic impact, social effects, technical requirements, legal implications, and environmental impact. For the purposes of this research, the focus falls mainly on the technical aspect.

7 Framework

The roles defined in the previous section are only viable if performed within the context of a larger, all-encompassing framework. This framework will consist of five facets, namely, stakeholder engagement; the building of necessary ICT infrastructure; the creation of new business models; awareness training and policy support; and monitoring and evaluation exercises. These are further explained below:

- **Stakeholder engagement**—Stakeholders are one of the key factors of this framework. A far from exhaustive list of this subset would include: *Learners, Teachers, School Administrations, Publishers, and Content Providers*. These stakeholders would need to come together to forge a path forward.
- **ICT infrastructure**—This consists of two parts: the physical and the non-physical aspects. The physical aspect entails structures such as adequate electricity grid coverage and ubiquitous internet connection. The non-physical aspect deals with software for the e-readers, cloud storage for schools, and creating language-localized educational material.
- **New business models**—ICT investments are capital-intensive, and it is necessary to have an accurate grasp of estimated costs and expected returns on investments (ROIs) in order to ensure sustainable spending practices. Key stakeholders in this subset would include government bodies (most notably the Ministry of Education), as well as any other entities or corporations providing development loans and subsidies.
- **Awareness, training and policy support**—It is necessary for new ICT implementations to be introduced together with sufficient training to enhance overall integration. This can be achieved through training workshops, seminars, and advanced professional development. Furthermore, supporting policies must be enacted to enable and raise awareness of the overall process.
- **Monitoring and evaluation**—After implementation, monitoring and stakeholder feedback solicitation are crucial in determining the success or ineffectiveness of e-reader deployment. These are needed to gauge the effectiveness and efficiency of implemented solutions against long-term goals.

8 Prototyping the Ideal Device

Based on the information gleaned from the data collection, analysis and role designations above, it becomes easier to imagine the design and capabilities of the ideal e-reader (Table 1).

The creation of apps for the device can be outsourced to local software developers within the country itself, thus building capacity, creating employment, and engaging more stakeholders in the transformation process.

Table 1 Attributes of the ideal e-reader

Physical attributes	Large screen, lightweight, durable, and scratch/shock resistant
Hardware attributes	• Long battery life for extended use • WIFI connectivity • Bluetooth/NFC (near field communication) • Touch screen for faster content creation using a stylus or finger • Camera
Software attributes	• Open source operating system • Built-in educational apps • Classroom management software for teachers • Ability for learners to “tag” content for later reference • Compatibility with several file formats and other devices (e.g., via USB connectivity) • E-Book search capabilities
Other attributes	• Low and affordable cost

9 Conclusion

9.1 Recommendations

The major recommendations resulting from this research are as follows:

- All core and auxiliary education stakeholders need to come together as an entity in order to map a way forward
- All key aspects of the implementation such as political, economic, social, technological, legal, and environmental facets of e-reader implementation should be addressed
- New business models need to be developed in order to cater to the proliferation of new ways of commercializing digital educational content
- Benchmarking exercises against the ICT implementation plans of other countries needs to be carried out in order to maintain high standards of efficacy and effectiveness. The results of these benchmarking exercises will need to be pitted against the realities of the Namibian context
- New policies will need to be formulated that support, encourage and reward innovative ICT implementation strategies within the country
- New teaching approaches will need to be developed and imprinted upon teachers in order to take full advantage of new pedagogical methods.

Future research in this area of study will involve further feedback solicitation from an increased sample pool of respondents. This will then be followed by the purchase of a few e-readers and small-scale piloting of them in selected schools to gauge the response, usability and ultimately the viability of implementing e-readers in Namibian schools. Further analysis will be required to determine the specific curricular needs of teachers and learners, and what can be done to improve upon the mistakes and challenges of the initial pilot.

The data collected from this research has shown that while there is a huge interest in going digital (64 % of students and 91 % of teachers surveyed positively supported the introduction of e-readers), the country still lacks the basic social, economic, and academic foundation needed to sustain relatively advanced initiatives like e-reader deployment in institutes of learning on a mass scale.

E-reader deployment in Namibian schools, however, would make an impact in several ways, including the creation of targeted educational material; the eliminated need to carry large, heavy books around; easier aggregation of learning material; easier updating of educational material; improved future acumen for similar devices; improved standardized testing; higher educational investment from poor rural parents; and availing longer study periods. Risks include theft, possible reduced profits from publishers, and isolation between learners as everybody becomes engrossed in their own device. With careful planning, however, these could be greatly minimized or avoided entirely.

References

1. Fischer, G.: The Namibian Educational System., Windhoek (2010).
2. Angula, N., Mutorwa, J.: ICT Policy for Education., Windhoek (2005).
3. Katire, F., Kavirindi, F.: Guide to ICT Initiatives and Research Priorities in IST-Africa Partner Countries. (2012).
4. ETSIP: The Strategic Plan For The Education and Training Sector Improvement Programme (ETSIP): 2005–2020. Strategic Policy Plan, ETSIP, Windhoek (2005).
5. Government of the Republic of Namibia: Namibia Education For All - National Plan of Action 2002–2015., Windhoek (2002).
6. Namibian Sun: ICT sector must be smart - Simataa. In: Namibian Sun. (Accessed September 22, 2013) Available at: <http://sun.com.na/education/ict-sector-must-be-smart-simataa.57544>.
7. Reeves, T., Herrington, J., Oliver, R.: A Development Research Agenda for Online Collaborative Learning. *Educational Technology, Research and Development* 54(4), 53–66 (2004).
8. van den Akker, J.: Principles and Methods of Development Research. In van den Akker, J., Branch, R., Gustafson, K., Nieveen, N., Plomp, T., eds.: *Design approaches and tools in education and training.*, Dordrecht (1999) 44–58.
9. Herrington, J., McKenney, S., Reeves, T., Oliver, R.: Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In Montgomerie, C., Seale, J., eds.: *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, Chesapeake, VA, pp. 4089–4097 (2007).
10. Reeves, T.: Enhancing the Worth of Instructional Technology Research through “Design Experiments” and Other Development Research Strategies. In: *Annual Meeting of the American Educational Research Association* (2000).
11. Bannan-Ritland, B.: The role of design in research: The integrative learning design framework. *Educational Researcher* 32(1), 21–24 (2003).
12. Gay, L. R.: *Educational research: Competencies for analysis and application* 4th edn. Merrill, New York (1992).
13. Barab, S., Squire, K.: Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences* 13(1), 1–14 (2004).
14. Cobb, P., Confrey, J., diSessa, A., Lehrer, R., Schauble, L.: Design Experiments in Educational Research. *Educational Researcher* 32(1), 9–13 (2003).

15. van den Akker, J., Gravemeijer, K., McKenney, S., Nieveen, N.: Introducing educational design research. In van den Akker, J., Gravemeijer, K., McKenney, S., Nieveen, N., eds.: Educational design research. Routledge, London (2006) 3–7.
16. Glaser, B., Strauss, A.: The discovery of grounded theory: Strategies for qualitative research. Aldine (1967).

Multi-region Pre-routing in Large Scale Mobile Ad Hoc Networks (MRPR)

Majid Ahmad Charoo and Durgesh Kumar Mishra

Abstract A wireless mobile ad hoc network has high node mobility and therefore requires dynamic schemes to address the process of routing. Periodic updates of the varying layout of a given wireless mobile ad hoc network may provide a solution. The frequent broadcasting of this routing information to all the participating nodes will incur overhead on rare resources of bandwidth and battery power of network participating nodes. The contribution of this work is the extension of our previous work to decompose a given large scale wireless mobile ad hoc network into various network regions thus limiting the broadcasting of network routing information within a region. Further a novel multi-region routing scheme is introduced for the process of path finding between a sender and receiver. A multi-region routing scheme will route the data transfer from one network region to another network region.

Keywords Multi-level routing • Large scale networks • Dynamic routing

1 Introduction

Routing is one of the basic network functionalities. Routing in wireless communication has received a lot of attention in past because routing has always been perceived as a bottleneck for overall network performance. The effectiveness of routing protocols directly affects network scalability, efficiency, and reliability. With continuing growth of wireless network's sizes, it is significant to develop routing protocols that not only achieve the basic design goals, but also cost minimum routing time.

M.A. Charoo (✉)

Department of Computer Science & System Studies, Mewar University,
Chittorgarh, India
e-mail: majidcharoo@gmail.com

D.K. Mishra

Department of Computer Science & Engineering, SAIT, Indore, India
e-mail: durgesh.mishra@sait.ac.in

In infrastructure-based cellular wireless networks, access points act as centralized centers for network management and control. The access points help not only in route decision-making but also act as fully equipped routers. The decision of routing path selection becomes far easy as participating nodes greedily select the strongest available access point in terms of signal strength. Routing paths in cellular networks are mostly static or may seldom change. The reason for static routing is the less probability for cell change or handoff. Thus, the process of routing for infrastructure-based wireless networks is uncomplicated and undemanding. An ad hoc network is a distributed group of nodes without a centralized control, where each node independently accedes to this network for communication. These independently existing nodes work on a multiple-hop data transfer model. A node may act as a sender, a receiver, or an intermediate node. The role of these nodes may change frequently as a result of high network mobility and change in node size. The role of intermediate nodes in a wireless mobile ad hoc network can be more than a simple participating node. The process of routing in ad hoc wireless networks is quite different from the infrastructure-based wireless networks. As node mobility factor is high, the absence of centralized control center adds to the complicacy of routing process. An ad hoc network changes its topology dynamically as a result of reasons like, high node mobility, addition of new nodes to the network, and relinquishing of participating nodes. The change in topology requires change in path selection for data transfer from source to destination node. A large scale mobile ad hoc network consists of an arbitrarily large number of nodes (hundreds or thousands), randomly deployed in an outdoor area. The nodes work together to send the data to an intended destination node. The routing process can be based on proactive or reactive routing scheme approaches. In proactive routing, the network topological information is maintained at every node. Such a strategy avoids the need for establishing routes for each message and is efficient when the network topology is relatively static and traffic is relatively heavy. Reactive routing, on the other hand, does not maintain global topological information. When a message arrives, the source floods a request packet over the network searching for the destination. Such a strategy avoids the need for frequent topological updates and, therefore, substantially reduces periodic network updates.

In this paper, we consider a theoretical approach to analyze routing strategy of general wireless ad hoc networks in which links are subject to random breakdowns and network topology varies in time. The remaining paper is structured into various sections as follows. Section 2 provides a detailed survey of past and noteworthy research work related to wireless data routing and concludes with the main research gaps in the related area. Section 3 introduces the multi-regional network decomposition for wireless mobile ad hoc networks. Section 4 demonstrates the results with the discussion derived from the introduced scheme.

2 Related Work

An ad hoc network is a distributed group of nodes without a centralized control, where each node independently accedes to this network for communication. These independently existing nodes work on a multiple-hop data transfer model. The data transfer commences only after the route from source to destination is defined. The primary goal of any ad hoc network routing protocol is the establishment of correct and efficient route between a pair of nodes so that data may be delivered in a timely manner. The past research in the area of routing in wireless mobile ad hoc networks has been based on the routing schemes adapted from the already existing schemes present in general wired and wireless networks [1]. In recent years, several routing protocols have been proposed for mobile ad hoc networks and prominent among them are DSR, AODV, and TORA [2, 3]. Several adaptations in standard routing schemes have been introduced in the wireless standards [4]. The usage and applications of data communication among the nodes of wireless mobile ad hoc network kept on changing, so the quality of service became a concern with the advent of ad hoc wireless technology [5, 6]. The aspect of power backup as still is an issue of great concern. Routing overhead has an impact on the consumption of this scant resource. Lot of work in the area of sensitive routing schemes came into front taking battery consumption as an issue of prime concern [7, 8]. Much research has been done in recent years investigating different aspects like low power protocols, network establishments, routing protocol, and coverage problems of wireless ad hoc network [9].

The recent research in the area of critical link identification within a network has improved the quality of wireless network design by allowing logical network decomposition. The flow of network data traffic can be managed efficiently if the location of critical link is known in advance [10]. Different techniques for critical node detection have been devised for wireless networks in general and ad hoc networks in particular [10–12]. The author has devised an efficient critical node detection technique for network decomposition [13]. The future of wireless technology is going to be realized as large scale distributed networks, where challenges related to routing are going to be of prime concern. The routing of data in these large-sized networks spanning over spatially distributed regions has to be analyzed and proper routing schemes are to be developed. The existing schemes of routing, especially in wireless mobile ad hoc networks, inherit a bottleneck of node number. These existing schemes are obsolete to cater with the large-sized wireless mobile ad hoc networks.

3 Multi-region Pre-routing (MRPR)

The complexity of route calculation in a mobile wireless mobile ad hoc network is a non-polynomial problem. Routing protocols generally use either distance-vector or link-state routing algorithms to find the shortest paths to destination. The shortest

path finding algorithms always induce complexity to the routing process. This process becomes more complex for large scale mobile ad hoc networks. Another issue related to routing is the bandwidth consumption by routing packets for path updating and route discovery across the network in proactive and reactive routing schemes, respectively. The routing overhead depends on the number of nodes in the network and network mobility. In large networks, the transmission of routing information will ultimately consume most of the bandwidth and consequently block applications rendering it unfeasible for bandwidth limited wireless ad hoc network. In a network with N nodes, link state updating generates routing overhead on the order of $O(N^2)$. This order of routing overhead is the major issue in large scale ad hoc networks. Routing complexity due to reasons of bandwidth and complexity has an on the average battery power consumption of nodes across a given network. In a large scale wireless mobile ad hoc network the packets for route discovery and route updates enhance the battery consumption drastically. Thus reducing routing control overhead becomes a key issue in achieving routing scalability. A large sized mobile ad hoc network may be decomposed logically into smaller subnets to achieve better efficiency in complexity bandwidth and energy issues. The next section of this paper describes the logical decomposition of a large-sized mobile ad hoc network.

3.1 Multi-region Network Decomposition

A mobile ad hoc network as usually realized by an undirected graph $G(N, E)$ comprising of mobile wireless nodes (stations) denoted as 'N' and a total permutation of connections (links) denoted as 'E'. The operation state (working/faulted) of a link at any point of time may define the configuration of the network. Here we define a critical link as the link which provide only means of communication between different parts of a network (the removal of the link will disconnect the communication from source to receiver). A critical link divides a network into regions or subnetworks of smaller size. These regions of network nodes are supposed to be connected via the critical links. The probability of presence of critical nodes can be justified by our previous work [13]. The random presence of critical nodes in a network is shown in Fig. 1. The division of a network into a number of subnetworks thus reduces the complexity radically. The present proposed solution for the problem is based on subdividing a large mobile ad hoc network into possible subnetworks. As evident the decomposition of a large network into small subnetworks will ultimately have an effect on the actual problem of routing overhead reduction. This reduction of routing overhead will in turn reduce the bandwidth consumption. Thus allowing actual applications/data consume the maximum bandwidth. Figure 2 describes a general large scale mobile network. In this figure

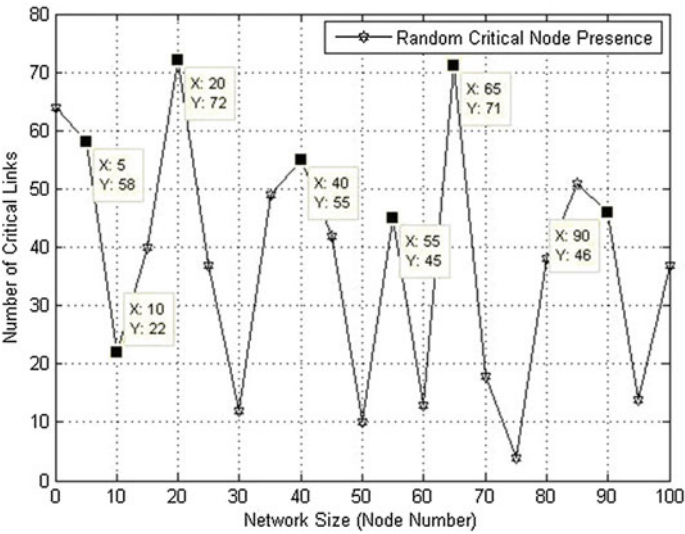


Fig. 1 Presence of critical nodes in wireless mobile networks

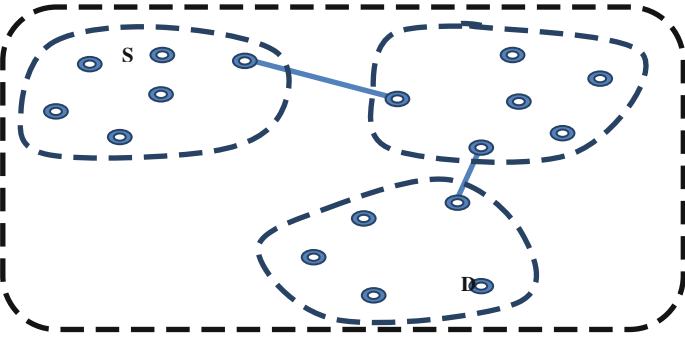


Fig. 2 A large scale wireless mobile ad hoc network

we can see small islands of a network distributed in a nonuniform manner. The islands are connected by a fewer connecting links. These links are the only means of communication between the network islands. Here the presence of these links is exploited for logical network decomposition. The present standard routing schemes consider all the nodes between source and destination. A pre-routing process can be done in order to decompose the network before actual routing scheme is applied. The complete routing process can be defined below.

3.2 MRPR Algorithm

The implementation of MRPR protocol can be divided into three phases.

Start-of-MRPR Algorithm

Phase 1. Network Decomposition

- Step 1. Define the intended source and destination nodes.
- Step 2. Using our critical node detection technique, identify critical nodes in the given wireless network [13].
- Step 3. Decompose the whole network into possible subnetworks.

Phase 2. Multi-Subnet Route Discovery

- Step 4. Broadcast route request RREQ messages swithin subnet sseeking for the destination.
- Step 5. For each (network subnet), Using AODV find route for data transfer.

Phase 3. Actual Data Transfer

- Step 6. Transfer the actual data over the multi-subnet route.

End-of-Algorithm.

4 Simulation Environment

The proposed protocol MPRP was simulated using NS2 network simulator. Table 1 shows the details of the simulation environment features.

Table 1 Simulation variables

Simulator	NS2
Routing protocols	AODV
Simulation area	1000 m * 1000 m
Various network sizes (node number)	25, 50, 100, 200
Transmission range (m)	250 m
Mobility model	Random way point
Maximum speed	5 m/s
Data packet size	512 kB
Traffic source CBR	Traffic source CBR
Maximum node energy	Joule

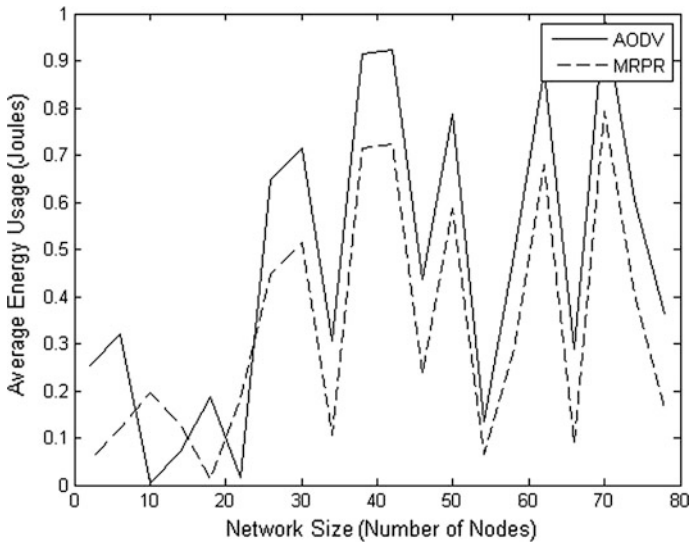


Fig. 3 Average energy consumption of simulated MANET

4.1 Average Energy Consumption

The study of batter power of network nodes in MRPR can be assessed by the comparison of the energy consumption of the nodes in basic AODV routing protocol as can be seen in Fig. 3. The proposed MPRP showed better results for the battery power as compared to AODV. The reason of improvement in the energy is due to the limited broadcasting of route discovery packets (RREQ). Also the amount energy required for route discovery is very less in large sized network gets restricted to only smaller subnets.

4.2 Average Routing Time

The study of routing overhead in efficient path discovery of network nodes in MRPR can be assessed by the comparison of the average routing time of the networks in basic AODV routing protocol as can be seen in Fig. 4. The proposed MPRP showed better results for the average routing time as compared to AODV. The reason of improvement in the route discovery time complexity is due to the limited configuration in a given subnet.

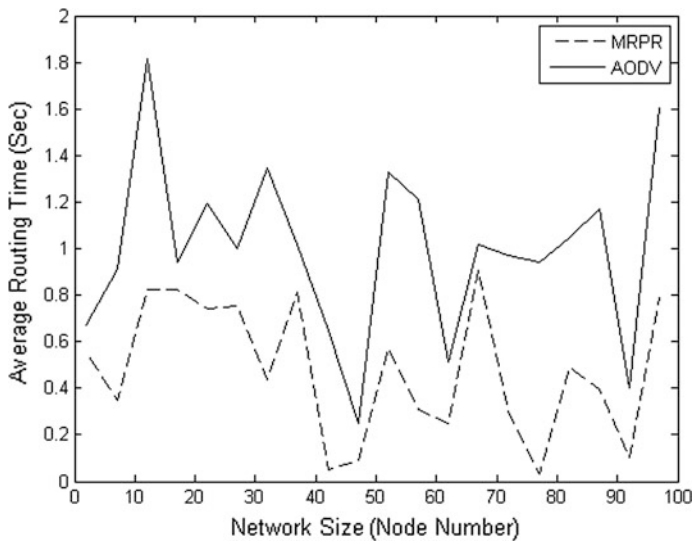


Fig. 4 Average routing time of simulated MANET

5 Conclusion

The introduction of MRPR, a pre-routing scheme for wireless mobile ad hoc networks, shows a clear improvement in the quality of service metrics. The metrics like average energy consumption and routing time has improved to a large extent. This work is an extension of our previous work in the area of critical node detection [13]. The current work will be taken to a more rigorous analysis and result generation to frame MRPR to be a more realistic scheme.

References

1. Johnson, David B., and David A. Maltz. "Dynamic source routing in ad hoc wireless networks." Kluwer International Series in Engineering and Computer Science (1996): 153–179.
2. Royer, Elizabeth M., and Chai-Keong Toh. "A review of current routing protocols for ad hoc mobile wireless networks." *Personal Communications*, IEEE 6.2 (1999): 46–55.
3. Karygiannis, A., E. Antonakakis, and A. Apostolopoulos. "Detecting critical nodes for MANET intrusion detection systems." *Security, Privacy and Trust in Pervasive and Ubiquitous Computing*, 2006. SecPerU 2006. Second International Workshop on. IEEE, 2006.
4. Taneja, Sunil, and Ashwani Kush. "A Survey of routing protocols in mobile ad hoc networks." *International Journal of Innovation, Management and Technology* 1.3 (2010): 2010–0248.
5. Singh, Shio Kumar, M. P. Singh, and D. K. Singh. "A survey of energy-efficient hierarchical cluster-based routing in wireless sensor networks." *International Journal of Advanced Networking and Application (IJANA)* 2.02 (2010): 570–580.

6. Sheng, Min, Jiandong Li, and Yan Shi. "Critical nodes detection in mobile ad hoc network." Advanced Information Networking and Applications, 2006. AINA 2006. 20th International Conference on. Vol. 2. IEEE, 2006.
7. Ben-Othman, Jalel, and Bashir Yahya. "Energy efficient and QoS based routing protocol for wireless sensor networks." *Journal of Parallel and Distributed Computing* 70.8 (2010): 849–857.
8. Srinivas, Anand, and Eytan Modiano. "Minimum energy disjoint path routing in wireless ad-hoc networks." *Proceedings of the 9th annual international conference on Mobile computing and networking*. ACM, 2003.
9. Cobo, Luis, Alejandro Quintero, and Samuel Pierre. "Ant-based routing for wireless multimedia sensor networks using multiple QoS metrics." *Computer networks* 54.17 (2010): 2991–3010.
10. M. Jorgic, I. Stojmenovic, M. Hauspie, and D. Simplot-Ryl. Localized algorithms for detection of critical nodes and links for connectivity in ad hoc networks. 3rd IFIP MED-HOC-NET Workshop, 2004.
11. Michaël Hauspie. Localized Algorithms for Detection of Critical Nodes and Links for Connectivity in Ad hoc Networks, in: "Proc. 3rd IFIP Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET 2004), Bodrum, Turkey", 2004.
12. A. Arulselvan, C. W. Commander, L. Eleftheriadou, and P. M. Pardalos. Detecting critical nodes in sparse graphs. *Comput. Oper. Res.*, 36(7):2193–2200, 2009.
13. Ahmad, M. and Mishra, D.K., "Critical Node Detection in Large Scale Mobile Ad hoc Networks", *International Journal of Computer Applications*, Vol. 82, No 17, PP 34–38, 2013.

Challenges Faced in Deployment of e-Learning Models in India

**Kamal Kumar Sethi, Praveen Bhanodia, Durgesh Kumar Mishra,
Monika Badjatya and Chandra Prakash Gujar**

Abstract IT industry is more demanding in technical skills which itself is vast and dynamic in nature; every day a new technology arises. A model developed for IT engineering students is discussed here which is an e-learning platform and facilitate the aspirant round the clock to nurture programming skills online. The developed system is a self review model where student can check the levels of his skills set and keep performance record for further references. This system provides hands-on practice experience and is accessible via any computational device which is Internet-enabled. This model will definitely help the conventional technical teaching pedagogy and groom the aspirants to be deployed directly in industry. In this paper, we also focused about the challenges and issues related to the e-learning model in the country. These issues could be used as a theoretical foundation to facilitate decision-making for the adoption of e-learning model.

Keywords e-learning · MOOC · Online education · IT in education · Collaborative learning

K.K. Sethi (✉) · P. Bhanodia

Acropolis Institute of Technology & Research, Indore, India
e-mail: kamalksethi@gmail.com

P. Bhanodia

e-mail: pcst.praveen@gmail.com

D.K. Mishra

Sri Aurobindo Institute of Technology, Indore, India
e-mail: durgeshmishra@ieee.org

M. Badjatya

Acropolis Faculty of Management & Research, Indore, India
e-mail: monikasethi1207@gmail.com

C.P. Gujar

Mahatma Gandhi Chitrakoot Gramodaya Vishwavidyalaya, Chitrakoot, India
e-mail: cpgujar69@gmail.com

© Springer Science+Business Media Singapore 2016

S.C. Satapathy et al. (eds.), *Proceedings of the International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 438, DOI 10.1007/978-981-10-0767-5_67

1 Introduction

With the advancement in information technology, the problem of accessing technical education has been resolved and open to all to grab a certification course of his interest for boosting career. As the technology is budding like anything, so are the requirements of the qualified and skilled people which will play a remarkable role in pushing the organization up and the progress of the nation. Technology has brought many opportunities in the area of learning and education at the doorstep with the help of Internet in the form of e-learning models or virtual training programs. In today's knowledge-driven economy, education has become more important than ever before. In knowledge-based economy, globally job assignments demands sharp technical skilled executives. This expectation is not only from fresh graduate but also from experienced employee, which force the employee to undergo for advance technical skill sets. But, due to social and financial responsibility they are not able to undergo traditional university program which limits their further promotions. Technically skilled employees are more productive and receive higher wages. Since to access the technical skill round the clock is quite cumbersome and consequently the people are less skilled, either due shortage of potential technical trainers or because of some other reason due to which the resource is inaccessible. Online training provides unique opportunities which are quite capable to overcome these problems. With the help of online technical training program, variety of technical skill sets is more accessible and available at any place with the help of Internet.

Even after there are many challenges in present status of e-learning systems; however, by overcoming these challenges we can have a great system to train the dumb. Information technology has become the key to a new world of online education. The e-learning model has become one of the most popular ways of gaining access to higher education for adult and professionals. In this IT era, technology is changing rapidly and today's world we live is much different than the one we live just 20 years ago. This rapid change has forced everyone to change and cope with the expected demand. Education institutions are also trying to cope up with these technological changes but still they are busy in making money. The e-learning offers new possibilities to all potential students and who cannot go to IT finishing school; now the IT finishing school could be sent to them. Online education is facing biggest challenge due to the New Digital Divide. For the development of global knowledge economy, the role of ICT in fostering online e-learning model is discussed by Elias et al. [1]. In online education, the digital divide defined as the gap between those learners who have, do not have, and know how to use the Internet and the information technologies. Tyler and Alex illustrated that online education is more flexible and economic over traditional teaching and these advantages will grow with improvements in information technology (IT) [2]. Clark and Mayer have defined e-Learning as instruction delivered by any technological mode intended to promote learning online [3]. According to Bilal, learning is self-paced in e-learning environment and students have choice to accelerate

according to his need. The e-learning model provides option to students to choose content and tools suitable to his interests, needs, and skill levels [4]. Birch and Burnett pointed out that in spite of the advances in e-Learning technologies and good practice, the adoption of e-learning model has failed to reach the predicted expectations [5]. Huong argued for students' preferred ways to learn can play an important role in adaptive e-Learning system [6]. Wang has suggested the need and importance of personalized assessment and content for learner in e-learning model which accelerate the adoption of e-learning and proposed GPAM-WATA e-learning system [7].

2 Advantages of Applying e-Learning Models

The e-learning education systems have many benefits over the traditional classroom teaching program. These benefits do vary with respect to student, faculty and institution. It provides convenience and flexibility to teacher and student both. Online teaching offers more options to teacher for engaging instructional activities. Teacher can teach from anywhere, any time in flexible mode. The e-learning provides opportunity to know your students better. An e-learning system provides a comfortable environment to students to participate in course activities and discussions. As the online education system is entirely computerized system, each activity of student is recorded, automatically critically evaluated and available for future reference. With this timely evaluation the trainer or teacher can keep an eye on student learning curve and can give timely feedback to improve further. In online education, learners' engagement and learning increases as all students are required to participate in discussion on the forum. To participate actively, every student has to work hard going through different kind of problems and coming up with new ideas and subsequent solutions.

The most important benefit of e-learning course is students may access to the resources online round the clock. With e-learning, one can complete his course as per his convenience and get certified. Online education provides a fair opportunity for learning to adult, professional, working populations. Traditional educational system requires student engagement in specific time duration and will not be available after certain periods of time, whereas e-learning program are available as per need of students. Other benefit of e-learning is that it can be offered to mass without limiting the batch size which generally happened in traditional system. Students do his study as per his convenience and he feels that he is more creative he can write programs using any tool as discussed in the implemented case study for this e-learning model. Online teaching reaches to mass that would not have been possible in traditional system. The students have the choice to choose their own technology from the place he feels most suited.

The system would surely facilitate the challenged students by providing them easy access to the interface round the clock. As physically challenged students would struggle to navigate a physical campus, or may get hearted due to partial

treatment from peers. But online education can give them opportunity to complete their study without facing these aforesaid problems. Even in traditional system, it is very difficult for an instructor to teach students with different sensory disabilities. The e-learning programs are required to be carefully designed in such a way so that they can address demand of these physically challenged students. The e-learning courses should be created in such a way that all course material should be accessible in different ways, be it through audio or video or text.

Teachers find increased efficiency in some wrote tasks. Online teaching tools automate processes and save instructors time and drastically reduce the amount of time spent evaluation or grading. In e-learning courses, teachers feels more connected with their students and are able to get to know them better than they thought possible. Teachers get participation of students who rarely take part in class discussions are more likely to participate online. It is experienced by teachers that students of e-learning courses wrote better papers, performed better on exams, produced higher quality projects. Many teachers who used mixed approach of teaching report that the use of Blackboard has increased their efficiency because they organize the course online and automate some basic activities such as quizzes, grading, and announcements.

3 Application Challenges

Demand of e-learning system is very high and recent report by Sloan Consortium shows that enrollment in online education is at an all time high [8]. But still online education is facing many challenges including technology, technical, content development, student motivation, etc., which limit the adoption of e-learning. In traditional education system, teacher keeps their students motivated and gives them input time to time for study, work, etc. They ask students to come to college regularly and keep up their performance with their personal involvement. But in case of online education this kind of motivation is missing and because the online education comes with the philosophy learns anywhere anytime. So, it becomes difficult to force student to do study. Student works alone in a virtual environment in isolation and requires very high commitment from him to get motivated. Not only the motivation but self discipline is also a major point of concern. In online education, there is no set times for classes and no designated place for study. Due to philosophy of study any time students do not study regularly and get overburden with pending work. This freedom some time lead to procrastination and due to which he will not able to complete his assignments on time, or might not be able to cope with the work and in result many students left their course in between.

Developing countries like India, availability and adoption of technology is a big challenge. In India, still computer comes under luxury and the availability of Internet is subjective. To promote e-learning, we need to solve this problem and if we cannot offer computer and Internet to individual we can have community hall where this can be made available at nominal cost. The hardware and peripheral

devices are not everlasting and in case of breakdown they must be able to fix the problem. The delivery model must be made on technology which can support old platform too because in India like country people are not financially capable to buy all new technological development. If these points are not taken into consideration, it will not be able to work efficiently and any e-learning delivery model will fail in India which will lead frustration in student.

In e-learning education generally the communication made is written whether it is content delivery or submission of assignments. Due to which it students lack on oral communication skills and to improve on this they need to undergo traditional program. To work on this, in online education, we can promote assignments submission in the form of video presentation and can have online discussion between peers and teacher. Courses which require lab or hands-on training may not fulfill the purpose completely in online mode. Generally, labs are simulated and practical are offered online but due to technology constraint if it cannot be done then students has to go for lab work in traditional mode. Possibly the biggest challenge faced by e-learning could be adoption of technology by students. It may be possible that students are not tech savvy and this could be more with older adults. This problem may be solved or overcome by giving an additional short-term training on the use of technology and enhance student's basic technical skills. Before taking the e-learning course, students must consider these challenges and be prepared to overcome these challenges. If he feels any challenge seems insoluble, the student should give a thought before talking e-learning course and may opt traditional course if it suits.

Challenges are not only faced by students but there are several challenges which are faced by teacher and need to be answered for and efficient delivery of course. Archambault discussed that time needed to design and implement a structured online lesson is an important point to be considered and because of new content, new technologies normally the time required to create e-learning courses increased [9]. The biggest challenge faced by teacher in e-learning course is to built-up a community of learners. To build a sense of community between students is a challenge in an e-learning course. In traditional model a sense of belongingness is there and community develops naturally whereas in e-learning courses teachers have to take care of this. In e-learning course the mode of delivery is different than the traditional mode and teachers have to re-envision his curriculum, goal, content, assignment, activity, and evaluation method as per mode of delivery and it should be planned in such a way that student can get benefited at the maximum. Many time students need interaction with teacher and the learning model must be designed for this need. Online discussion forum, chat room and facility of video conferencing may facilitate this and give ease to teacher. However, the success of e-learning course depends upon the involvement of teacher and should be competent enough with the technology to participate actively in this.

4 Future of e-Learning—Massive Open Online Courses (MOOCs)

Due to the availability and adoption of online courses several million people are enrolling in these courses and completing their study and catering their knowledge needs. Online course makes students enable to choose their way from variety of providers, courses, credit transfer and credit assessment in order to lead to their chosen outcomes. It allows student to complete their course on fast track at affordable cost and credit rating become more ubiquitous. Massive Open Online Courses (MOOCs) is a recent development in the education field where many virtual universities such as Coursera, EdX, Udemy, and Udacity have taken shape. This technological shift has enabled millions of students from all over the globe to take humanities, management, science, and engineering courses along with their job, offered by the world's best professors at nominal fee or free. These MOOCs offer students to gain knowledge from courses offered by the world's top universities. Question that arises here is—Is this end of traditional education system or a new birth of New Ireland. It is not the substitute of traditional education system and major MOOC model have come from top notch university like Stanford, MIT, Harvard, etc. Even content for personal initiatives like Course are supplied by universities for free. But, the MOOC model is a troublesome for these universities, why would universities take these initiatives. In MOOC model, learner learns through the experiences of other remote peers and gets exposed to a variety of individuals from different demographical place.

The best feature of this model is that it has no limitation in batch size and any number of the student can join at any moment to start the desired course from anywhere round the globe. As student from different locations and culture it is an opportunity for them to improve their interpersonal and communication skills to enrich their course knowledge. MOOC gives opportunity to learner to join course of his interest from any place anytime however a sincerer commitment to pursue course is expected from the learner but in many cases this quotient is missing and too hard to overcome as this depends on the nature of individuals. This drawback puts an adverse affect on the students performance and consequently too hard to maintain the academic integrity standards. Due to the non-availability of law on intellectual property rights for open access models many times professor hesitates to contribute in this model.

5 The Issues Answered by Proposed Solution

The challenges faced during online could be eliminated by developing a model which eventually resolves the issues as under:

- The first and prime objective of any students belongs to CS/IT is programming skills and it can only be developed when on do it by own and the developed system would definitely provide space for hands-on practice with facility to run and compile.

- It is always difficult to identify a well equipped space where students can perform the lab activities easily; with the use of this interface it is easy for the students to access the system from anywhere and any moment of time.
- Evaluation of lab exercise altogether of 20 or more students at a time is fairly a difficult task from a single observer perspective because all of sudden he would not be in position to check everyone's program, the developed system provides the facility for evaluation of program line by line and instantaneously result could be seen.
- A hardware should be installed with necessary latest version of required setup of the tool in which the student can write program, hence every machine in the lab should be installed with the required software else it would be nearly impossible to write one program.
- There is always a virus threat possible on any machine; it could corrupt your done work anytime. As the system is online hence no such threat can harm the work done and the portal is not supposed to be corrupted from any sort of virus.
- Time-to-time, the high-end versions of the software are released and to run such software high end machines too required which is such a challenge that could be difficult for everyone to meet and there is no such up gradation is needed the only thing required is the access of Internet.
- In labs, students are bounded couple of hours and resources are inaccessible to them, eventually using this system students can be connected with the platform round the clock consequently they are in a position to work as and when required.
- The system is such devised that it can run on any hardware platform which is having Internet even on the latest gadgets and smart phones hence no need to carry laptops even.
- The system is quite powerful in keeping records and reports of programs can easily be taken.
- It will be capable enough to recall the programmer as how long he is keeping himself away from a keystroke.
- This interface provides bunch of questionnaire for assessment of individual.

Developed solution, xtremelearning is a powerful online course management system developed using PHP and Wordpress. The system facilitates multiple languages and offer options for creating online courses, lesson management, quiz management, content in audio and video, questions management, and tracking course and student progress, etc. Working with developed system is really very easy and increase teacher and learner satisfaction from online training in easier way.

Mobility is the key to success in today's world and developed system is compatible ranging from different platforms, browser and mobile device like Android and iPhone. The system has option to create courses, write lessons, and add quizzes to test your learners, set lesson and course pre-requisites, allow user registration and charge students for paid courses. You can build your course with minimal effort and reuse presentations and videos or other available materials. The system gives option to user to customize look and feel as per his needs.

Sign up process is simplified and quick user registration make learner feel comfortable. After getting log in system, learners have access to a personalized dashboard and can track his course progress. Provision for simple yet comprehensible analytics about everything that happens inside system is made and sensible reports are designed for different needs. Course analytics provide you with an overview of your content, grades, as well as the students who are registered with the system. Option to test student with a variety of question types is available and virtually has no limit to the kind of quizzes teacher can create. From the created question bank, questions would be display in random sequence to learners while taking test. The grading of the quiz can be set on teachers choice either automatic or manual depending upon the requirements. This system throws a wide range of different language online compilers embedded within it so that student can practice his programming skill then and there without any external support. Moreover, the interface also allows the learner to save his or her developed program in the login for further use.

6 Conclusion

Recent data analysis show that only 17 % of engineers produced in the country are employable rest could not even be counted. Great reason to think upon the quality of skills professionals is having. The e-learning system is an innovative holistic approach that provides an interface for learning that meets today generation's objectives in a convenient way at their doorstep with mobility. Such models are collaborative models and have array of choices to make learning experience more interesting. However, limitation of technology and usage of technology is a big question in implementing such models and must be addressed before offering or taking a course. Today's learners though are digital native proper use of available technologies for such models from both the sides would motivate the use of e-learning system approaches effectively. Perhaps, selecting and designing the course content is the key of success for any e-learning course.

Further, it could be concluded that the developed online learning system would definitely provide an interface for IT professional aspirants to sharp their skills and could be well worse in particular technology. With the help of e-learning education, we can cater to mass that is unmet and under-served like working professionals, rural and military population. Getting skilled in today's era is the one significant prerequisite for success in anyone's career. This paper has canvassed a proposed a technical e-learning portal for the people of information technology along with a brief critical review of e-learning education approaches that helps the students to decide which model of learning has to be chosen for their further higher education or so.

References

1. Elias, G.C., Denisa, P., Caroline, S., McDonald, S.: Technological learning for entrepreneurial development (TL4ED) in the knowledge economy (KE): Case studies and lessons Learned. *Technovation*, pp. 419–443, Elsevier doi:[10.1016/j.technovation.2005.04.003](https://doi.org/10.1016/j.technovation.2005.04.003) (2006).
2. Tyler, C., Alex, T.: The Industrial Organization of Online Education. *The American Economic Review*, vol. 104(5), pp. 519–522, American Economic Association, doi:[10.1257/aer.104.5.519](https://doi.org/10.1257/aer.104.5.519) (2014).
3. Clark, R., Mayer, R.E.: *E-learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning*, Pfeiffer, San Francisco (2011).
4. Bilal, S.: *E-Learning Revolutionise Education: An Exploratory study*. In: *E-learning: A Boom or Curse*, ISBN-978-93-80748-87-0 (2015).
5. Birch, D., Burnett, B.: Bringing academics on board: Encouraging institution wide diffusion of e-learning environments. In: *Australasian Journal of Educational Technology*, vol. 25(1), pp. 117–134 (2009).
6. Huong, M.T.: Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. In: *Computers in Human Behavior*, ISSN 0747-5632, Elsevier, doi:[10.1016/j.chb.2015.02.014](https://doi.org/10.1016/j.chb.2015.02.014) (2015).
7. Wang, T.H.: Developing an assessment centered e-Learning system for improving student learning effectiveness. *Computers & Education*, vol. 73, pp. 189–203, Elsevier, doi:[10.1016/j.compedu.2013.12.002](https://doi.org/10.1016/j.compedu.2013.12.002) (2014).
8. Survey of Online Learning 2014, <http://sloan-c.org/publications/survey/index.asp>.
9. Archambault, L.: Identifying and addressing teaching challenges in k-12 online environments. *Distance Learning*, vol. 7(2), pp. 13–17 (2010).

A Novel Cross-Layer Mechanism for Improving H.264/AVC Video Transmissions Over IEEE 802.11n WLANs

Lal Chand Bishnoi and Dharm Singh Jat

Abstract This paper proposes a Novel Cross-layer Mechanism (NCLM) for Improving H.264/AVC Video Transmissions over IEEE 802.11n wireless network. According to the network traffic loads and the importance of the video data, the proposed mechanism dynamically selects the suitable access categories (AC) instead of predefined AC. This proposed novel cross-layer mechanism gives the information about the importance of video packets to the MAC layer. Information about the network traffic load is available from the queue length of all access categories. During this research, we analyzed the performance of NCLM in both light and heavy load over IEEE 802.11n wireless networks. In this research, the performance of video traffic measured by average: Throughput, PSNR, VQM, and SSIM. Simulation results of this research demonstrate that the performance of proposed mechanism was higher in comparison to the results derived from IEEE 802.11n EDCA, CLOT, DACMM, IPB-Frame AMM, and Static Mapping algorithm.

Keywords QoS • Cross-layer mapping • Multimedia transmission • Video over 802.11n • H.264/AVC

1 Introduction

Nowadays, the IP video traffic over the wireless network is continuously growing due to the advances in wireless network technologies and smart mobile communication devices. According to Cisco's Visual Networking Index, it is estimated that

L.C. Bishnoi (✉)
Uttarakhand Technical University, Dehradun, India
e-mail: bishnoi@yahoo.com

D.S. Jat
Namibia University of Science and Technology, Windhoek, Namibia
e-mail: dsingh@polytechnic.edu.na

globally IP video traffic will be 80 % of all IP traffic business as well as a consumer by 2019, it was 67 % in 2014 in which video exchanged through peer-to-peer (P2P) file sharing is not included. In 2019, the global consumer traffic will reach the range between 80 and 90 % for all types of videos, i.e., IP video, IP-TV, video on demand (VoD), the Internet, and P2P. In 2014, it accounted that 54 % IP traffic on wired electronics equipment. However, in 2016, it is estimated that the IP network traffic on wireless and mobile equipment may be increased rather than traffic on wired equipment [1].

Today's IEEE 802.11n wireless networks can manage needs of critical wireless communication. IEEE 802.11n than 802.11a/b/g networks can increase six-time wireless networks performance. This fulfills the needs of modern, reliable multimedia communication and business-critical applications. With IEEE 802.11n technology, educational institutions, organizational, educational, and research networks can gain high reliability and more throughput than with the IEEE 802.11a/b/g networks. This technology also provides reliable wireless network connectivity for mobile users that the various range of mobility applications without compromising whole network performance. The data rates of IEEE 802.11n wireless connection consistently reached up to 300 Mbps, and it translates throughput of 185 Mbps for a sustained period.

The networked video, also called IP (Internet Protocol) video or video over IP, can have substantial benefits for educational and research networks, healthcare organization and small and medium-sized business. Before the IEEE 802.11n wireless technology, mobile diagnostic services in healthcare organization can be possible only through high-definition (HD) stream video over IEEE 802.11b/g wireless networks. Even though, an IEEE 802.11g wireless networks would not be reliable communication networks for high-definition streaming video. IEEE 802.11n fulfills the requirement of throughput rate for mobile diagnostic services provided by a healthcare organization and another bandwidth hungry video-streaming application.

It is estimated that in 2019, all devices connected to IP networks increased three times the global population. Therefore, for bandwidth hungry application like IP video, more research needs to be done for reliable video transmission. On wired as well as wireless networks, the Medium Access Control (MAC) plays the important role in the performance of end-to-end video transmission. MAC handles allocating the resources to the different type of applications or wireless stations. Many types of research conducted for video quality measurement on network and application layer. However, some of the research has been done for MPEG4 video traffic from the MAC layer perspective. Very few researches have been done for H.264/AVC video traffic from the MAC layer perspective.

The effect of various MAC-level parameters for reliable transmission of video over IEEE 802.11n wireless networks analyzed [2]. Subjective tests to relate

MAC-level parameters for received different types of video traffic performed. If MAC-level parameters used carefully, then this improved the H.264/AVC video quality over IEEE 802.11n wireless networks as this study show.

2 H.264/AVC Video Sequences

H.264/AVC is a block-oriented video compression standard based on motion compensation. It also known as MPEG-4 AVC (Advanced Video Coding). Its coded picture consists of some macroblocks that organized into slices. The standard divided into two main layers. The first layer is a video coding layer (VCL). It specifies motion compensation; transform coding, and entropy coding detail of the video encoding engine. The second layer is a network abstract layer (NAL). It encloses coded slices into the network object in the network. H.264/AVC consists a basic coding block is a macroblock that encoded in intra or inter mode. The Video frames coded into one or more slices. These slices have many fixed-size macro blocks. Because of its self-contain minimal decodable information is the slice decoding performs independently.

It supports five slices for coding types, known as I, P, B, SP, and SI. With the exception of reference pictures I, P, and B are similar to previous coding standards. I slices have intra macroblocks, and P slices have intra macroblocks with referencing inter macroblock. B slices have inter and intra macroblocks with referencing another macroblock. The higher compression ratio obtained by using P slice and B slice with referencing other macroblocks. SP stand for switching P and SI stand for switching I. The SP slice used for switching between P slices of same video. The SI slices work for switching randomly and also for recovering errors [3].

3 IEEE 802.11n

IEEE 802.11n technology was formally released in late 2009. It provides enhanced performance over prior IEEE 802.11 technologies. It operates at 2.4 or 5 GHz space for maintaining backward compatibility with prior IEEE 802.11 technologies. For improving MAC efficiency and channel utilization, the overhead minimized by using aggregation mechanisms in IEEE 802.11n. The aggregation technique supports a large number of frames transmitted together over the wireless network into a single aggregated packet. The aggregation method accomplishes higher system gain and useful for applications that have shorter packets size. Such applications are Voice over IP (VoIP). Two possible methods can use in IEEE 802.11n for meeting

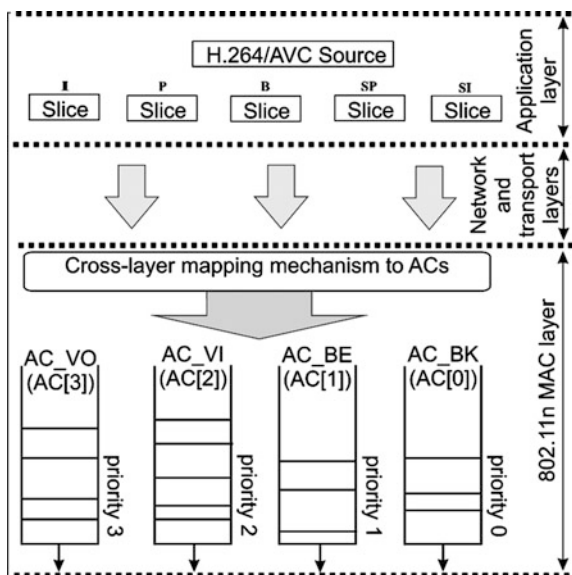


Fig. 1 IEEE 802.11n EDCA Architecture for H.264/AVC cross-layer mechanism

the requirements of higher throughput. First, one is enhancing the data rate of the Physical layer (PHY), and other one increases the efficiency of the medium access (MAC) layer.

The enhanced distributed channel access (EDCA) mechanism is a universal form of modified MAC. It is an enhancement of IEEE 802.11 original version standard's distributed coordination function (DCF). The IEEE 802.11n EDCA categorizes the traffic into four different access categories (ACs). Its service differentiation realized through four ACs at each station (STA). To make EDCA notifications simple we use AC_BK (AC for Background) as AC0, AC_BE (AC Best-Effort) as AC1, AC_VI (AC for Video) as AC2, and AC_VO (AC for Voice) as AC3 throughout this paper [4].

4 Cross-Layer Mapping

In cross-layer mapping mechanism, runtime allocates the video data to the most suitable AC according to the algorithm decision instead of providing a fix ACs. The IEEE 802.11n EDCA architecture for H.264/AVC cross-layer mechanism shows in Fig. 1. This architecture used for cross-layer mapping mechanism in which all traffic type queued in separate AC, instead of the entire traffic shared through a common queue as in DCF. It provides quality of service (QoS) for video transmission.

5 Existing Related Works

In this paper, we analyze five earlier related works for video transmission over wireless local area network (WLAN) with our proposed novel cross-layer mechanism (NCLM) work. We apply five algorithms of related work on IEEE 802.11n for H.264/AVC video.

IEEE 802.11n EDCA devices are backward compatible with IEEE 802.11 legacy devices. Features of IEEE 802.11e EDCA is also applicable for IEEE 802.11n. As in original IEEE 802.11n EDCA architecture, it sends packets to different ACs according to the incoming traffic category. Figure 1 shows four access categories of IEEE 802.11n EDCA are AC3, AC2, AC1, AC0 used for voice, video, best-effort, and background, respectively. The AC3 is the highest priority, and AC0 is a lowest priority access category [4].

Static mapping (SM) is a cross-layer mapping in which video slices I, P, B sends to AC3, AC2, AC1, respectively, and non-video traffic sent to AC0. The ACs accesses the channel as per their priority.

Cross-layer Optimum Techniques (CLOT) monitor the average queue length (Q_{avg}) by using current queue length, minimum threshold ($Threshold_{min}$), and maximum threshold ($Threshold_{max}$). If Q_{avg} is less than $Threshold_{min}$ the packets directly sends to queue. If Q_{avg} is larger than $Threshold_{max}$ then the packets are dropped. It calculates packet dropping probability which range from 0 to 1 [5].

Dynamic adaptive cross-layer mapping mechanism (DACMM) sends the slices as per the dropping probability of the slice's time. Initially, all slices of video sent to AC2, and slices moved into AC1 and AC0 as per the available space in AC2. New probabilities calculated according to the current queue length, a number of packets and threshold values [6].

IPB-frame Adaptive Mapping Mechanism (AMM) proposed constructing the relationship between video frames and voice access category AC3. It mainly focused on two video slices I and P. The limitation of buffer size of each AC queue the new I and P slice dropped for removing the congestion. For mapping, I slice on AC3 and mapping P slice on AC3 probability P_I and P_P calculated [7]. When higher priority AC3 is extremely busy, then P and B slices are sent to AC1 and decision-making done by calculating the probability P_P and P_B [7].

6 Proposed Novel Cross-Layer Mechanism (NCLM)

Algorithm 1 Proposed Novel Cross-layer Mechanism (NCLM)

```

BEGIN
  set  $Threshold_{low} \leftarrow 20\% \text{ of } AC \text{ Queue Length}$ 
  set  $Threshold_{high} \leftarrow 80\% \text{ of } AC \text{ Queue Length}$ 

  if  $qlen[AC_2] < Threshold_{low}$  then
     $AC_2 \leftarrow \text{Video packet}$ 
  else if  $qlen[AC_2] < Threshold_{high}$  then
    if  $qlen[AC_3] < qlen[AC_2]$  and  $sliceType = 3$  then
       $AC_3 \leftarrow I \text{ slice}$ 
    else
       $AC_2 \leftarrow \text{Video packet}$ 
  else
    if  $sliceType = 3$  then
      if  $AC_3$  and  $AC_2$  are full then
         $AC_1 \leftarrow I \text{ slice}$ 
      else if  $qlen[AC_3] < qlen[AC_2]$  then
         $AC_3 \leftarrow I \text{ slice}$ 
      else
         $AC_2 \leftarrow I \text{ slice}$ 
    else if  $sliceType = 0$  then
       $AC_1 \leftarrow P \text{ slice}$ 
    else
       $AC_0 \leftarrow B, SP \text{ or } SI \text{ slice}$ 
END;

```

The proposed Novel Cross-layer Mechanism (NCLM) is shown in Algorithm 1. When a video packet arrives, at that time, queue length of AC2 is calculated and compared with $Threshold_{low}$ (20 % of AC Queue length) and $Threshold_{high}$ (80 % of AC Queue length).

In NCLM, if the queue length is less than $Threshold_{low}$, all slices of video data (I, P, or B) mapped to AC2. If queue length between $Threshold_{low}$ and $Threshold_{high}$, all slices of video data mapped to AC2. However, if queue length of AC3 is less than AC2 then I slice mapped to AC3. Similarly, if the queue length of AC2 is more than $Threshold_{high}$ then P and B slices (SP/SI slice in H.264/AVC) of video data mapped to AC1 and AC0, respectively. In this conditions, I slice mapped to AC1 if AC3 and AC2 are full otherwise I slice mapped to AC3 or AC2 whichever is less occupied.

7 Experimental Scenario

The framework for video transmission over the WLAN in NS2 on Fedora operating system integrated with Evalvid and IEEE 802.11n framework used for the simulation for this study [8, 9].

For simulations, this research work developed a virtual machine integrated with NS2 simulator embedded with IEEE 802.11n module. In this work, a network topology was created using static stations where each wireless station transmits all type of traffic to its paired static station. Data rate 1Mbps configured between two static stations. In addition to H.264/AVC video traffics, research also created 256 kbps FTP traffic as background traffic and 125kbps CBR data traffic in between 2 IEEE 802.11n static stations. Foreman YUV QCIF (176 × 144 pixels) video traffic used as source for this research work. It contains 142, 146, and 266 packets/slices for I, P, and B, respectively. The video packets of 1500 bytes and 512 kbp data rate for H.264/AVC video traffic are set before transmission. Packet size and other simulation parameters are shown in Table 1. In the experimental study, 50 packets queue size selected for all ACs. In addition to the background and best-effort traffic, 64 kbps CBR voice traffic also created on the sender site.

Figure 2 shows the simulation topology configuration used in this simulation study. The topology consists of H.264/AVC multimedia server that connects to an 802.11n Access Point (AP). An AP connects to a mobile node using 802.11n wireless network.

8 Results and Discussions

The results of the proposed Novel Cross-layer Mechanism (NCLM) examined with the similar existing work’s result, e.g., IEEE 802.11n EDCA, Static Mapping (SM), Cross-layer optimization techniques (CLOT), Dynamic Adaptive Cross-layer Mapping Mechanism (DACMM), and IPB-frame Adaptive Mapping Mechanism (AMM). This simulation study uses four different traffic scenarios, which includes different loads of traffic on four ACs such as voice (on AC3), Video (on AC2), UDP (on AC1) and TCP (on AC0) as shown in Table 2. It generated video randomly and transmitted over IEEE 802.11n during the entire simulation period. In this paper, we examined the received video quality parameter using the average: Throughput,

Table 1 Simulation parameter

	VoIP	Video	Best-efforts	Back-ground
Transport protocol	UDP	UDP	UDP	TCP
Access category	AC3	AC2	AC1	AC0
Packet size	160 byte	1500 byte	200 byte	512 byte
Sending rate	64 kbps	512 kbps	125 kbps	256 kbps

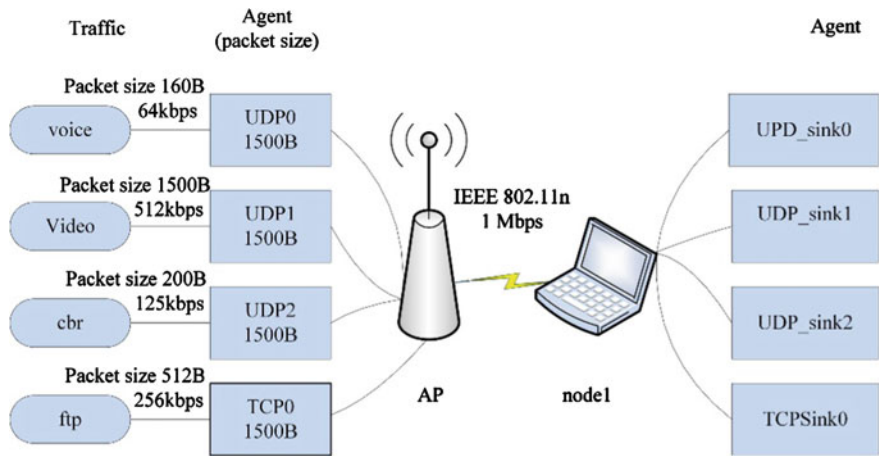


Fig. 2 Simulation topology

Table 2 Number of traffic stream in traffic scenario

	Voice AC3	Video AC2	TCP AC1	UDP AC0
Case 1	1	3	1	1
Case 2	5	3	5	5
Case 3	10	5	10	10
Case 4	10	10	10	10

PSNR, VQM, and SSIM for evaluating the effectiveness of proposed mechanism under various load conditions.

We also compared the loss of video slices for all existing methods with proposed NCLM method. Table 3 shows the number of H.264/AVC video slices lost during the transmission of the Foreman QCIF H.264/AVC source video. In heavy load, best mechanism saves essential I, P slices during transmission. The proposed mechanism NCLM save more I, P, B slices comparison than other existing methods. As shown in Table 4 proposed NCLM saves important slices I and P. In Case 1, there is no loss of slices. In Case 2, there is no loss in I slice and but only three P slices lost, which is less than another mapping approach. Similarly, NCLM saves I- and P-slices in heavy load condition during Case 3 and 4 also.

The quality measurement parameters for proposed NCLM compared with existing similar type of mechanisms shown in Table 4. The average throughput of EDCA 802.11n network under four different loading cases of H.264/AVC videos are shown in Fig. 3 and Table 4a. In all loading cases proposed NCLM gives higher average throughput than IEEE 802.11n EDCA, SM, CLOT, DACMM, and IPB-frame AMM approaches. Table 4a and Fig. 3 shows that static mapping (SM) average throughput almost equal to NCLM but other mechanism have

Table 3 Number of H.264/AVC video slice lost

Mapping type mechanism	Video traffic scenario											
	Case 1			Case 2			Case 3			Case 4		
	I	P	B/SP/SI	I	P	B/SP/SI	I	P	B/SP/SI	I	P	B/SP/SI
802.11n EDCA	0	0	0	4	9	14	72	64	109	108	90	166
SM	0	0	26	0	54	117	16	109	266	75	113	266
CLOT	0	0	0	2	11	22	31	67	143	96	106	212
DACMM	0	0	0	2	4	12	71	61	111	93	103	191
IPB-frame AMM	0	0	2	1	5	18	23	43	89	89	112	203
Proposed NCML	0	0	0	0	3	11	15	27	73	74	110	241

Table 4 Quality measurement of proposed NCLM with existing similar mechanisms

<i>(a) Quality measurement for average: Throughput and PSNR</i>									
Mapping type mechanism	Average throughput				Average PSNR				
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4	
802.11n EDCA	487.67	312.48	197.82	188.13	36.89	34.66	21.27	16.88	
SM	494.42	273.92	304.4	398.86	36.06	30.40	24.59	18.12	
CLOT	489.67	288.71	226.07	352.35	36.89	35.74	23.24	17.28	
DACMM	485.79	292.49	181.29	274.09	36.89	35.04	21.33	17.88	
IPB-frame AMM	490.38	294.17	218.24	281.2	36.89	34.52	21.68	16.06	
Proposed NCLM	494.67	312.71	304.96	444.3	36.89	35.84	29.16	18.52	
<i>(b) Quality measurement for average: VQM and SSIM</i>									
Mapping type mechanism	Average VQM				Average SSIM				
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4	
802.11n EDCA	0.96	1.91	7.26	10.28	0.96	0.91	0.58	0.44	
SM	0.96	1.80	7.04	9.48	0.96	0.93	0.59	0.44	
CLOT	0.96	1.07	7.00	9.22	0.96	0.95	0.64	0.47	
DACMM	0.96	1.35	7.20	9.22	0.96	0.94	0.63	0.46	
IPB-frame AMM	0.96	1.74	6.29	9.54	0.96	0.92	0.64	0.49	
Proposed NCLM	0.96	0.96	5.44	8.32	0.96	0.96	0.74	0.52	

average throughput lesser than NCLM. The proposed NCLM average throughput ranges from 304.96 to 494.67 Kbps.

Table 4a and Fig. 4 shows the average PSNR variations of transmitted H.264/AVC videos under four different loading cases. In Case 1 when the simulated network is light loaded, the proposed algorithm gives the almost similar 36.89 average PSNR as in other mechanisms. In Case 3 and 4, when network traffic has a heavy load then many slices lost as they moved into lower priority queues. The proposed NCLM mechanism dynamically handle the slices and gives better average

Fig. 3 Average throughput under four different loading cases

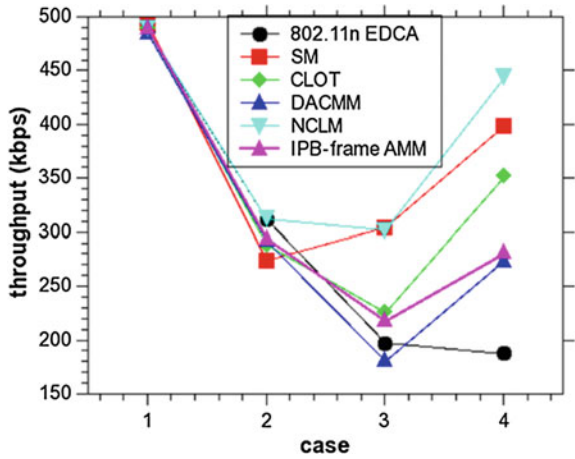
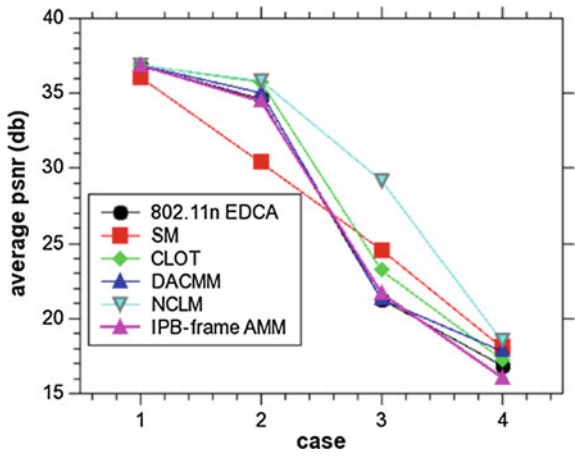


Fig. 4 Average PSNR under four different loading cases



PSNR comparison than IEEE 802.11n EDCA, SM, CLOT, DACMM, and IPB-frame AMM approach. The NCLM average PSNR ranges from 18.52 to 36.89 dB.

Table 4b and Fig. 5 shows the average VQM variations of transmitted H.264/AVC videos under four different loading cases. The smaller average VQM value shows better mechanism [10]. Therefore, proposed NCLM gives smaller average VQM values in all cases. The NCLM average VQM ranges from 0.96 to 8.32.

Table 4b and Fig. 6 shows the average SSIM variations of transmitted H.264/AVC videos under four different loading cases. The higher average SSIM shows better mechanism [10]. Therefore, proposed NCLM gives higher average SSIM in all cases. The NCLM average SSIM ranges from 0.52 to 0.96.

Fig. 5 Average VQM under four different loading cases

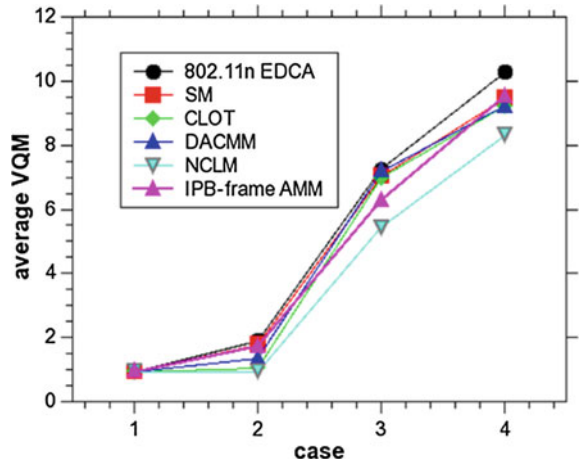
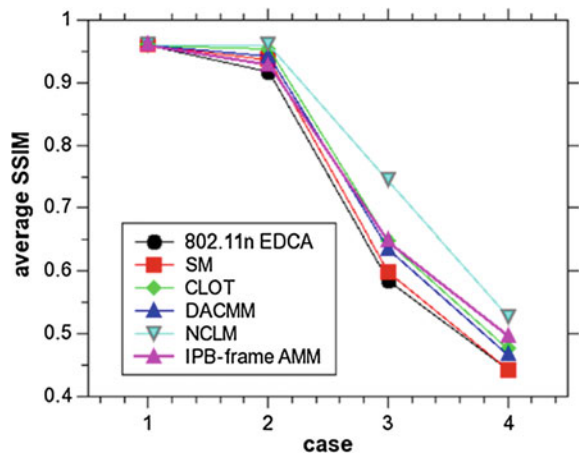


Fig. 6 Average SSIM under four different loading cases



9 Conclusion

This simulation analyzed the performance EDCA 802.11n for video transmission in light and heavy load using without mapping, static mapping, adaptive, and dynamic cross-layer mapping mechanisms. In this work, H.264/AVC video is used as a source for simulation. The average: throughput, PSNR, VQM, and SSIM calculated on IEEE 802.11n WLAN under four different loading cases. Results show that the proposed Novel Cross-layer Mechanism (NCLM) gives the higher average: throughput, PSNR, and SSIM value's comparison of other existing similar mechanisms, e.g., IEEE 802.11n EDCA, SM, CLOT, DACMM, and IPB-frame AMM. Similarly, NCLM also gives lesser average VQM values comparison of other existing similar mechanisms.

References

1. Cisco Visual Networking Index: Forecast and Methodology, 2014–2019 White Paper. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html.
2. Pokhrel Jeevan, Paudel Indira, Wehbi Bachar, Cavalli Ana, and Jouaber Badii: Performance evaluation of video transmission over 802.11n wireless network: A MAC layer perspective International Conference on Smart Communications in Network Technologies (SaCoNeT) 2014 pp. 1–6. IEEE, Vilanova i la Geltru (2014).
3. Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard, and Ajay Luthra: Overview of the H.264/AVC Video Coding Standard, IEEE Transactions on Circuits and Systems for Video Technology. 13, 560–576 (2003).
4. IEEE Std 802.11n-2009, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications”, in *Amendment 5: Enhancements for Higher Throughput*, ed, 2009.
5. Naveen Chilamkurti, Sherali Zeadally, Robin Soni, and Giovanni Giambene: Wireless multimedia delivery over 802.11e with cross-layer optimization techniques, Springer, Multimedia Tools and Applications. 47, 189–205 (2010).
6. Dharm Singh, Heena Rathore, and Chih-Heng Ke: Dynamic Adaptive Cross-layer mapping mechanism for video transmission over wireless networks International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)-2011, pp. 205–208. IEEE Explore, Udaipur (2011).
7. Xin-Wei Yao, Wan-Liang Wang, Shuang-Hua Yang, Yue-Feng Cen, Xiao-Min Yao, and Tie-Qiang Pan: IPB-frame Adaptive Mapping Mechanism for Video Transmission over IEEE 802.11e WLANs, ACM SIGCOMM Computer Communication Review. 44, 5–12 (2014).
8. Lal Chand Bishnoi, Dharm Singh, and Shailendra Mishra: Simulation of Video Transmission over Wireless IP Network in Fedora Environment, IP Multimedia Communications, A Special Issue from IJCA. 10, 9–14 (2011).
9. WMNLAB IEEE NS-2 802.11n Module. <http://wmnlab.ee.ntu.edu.tw/a/ns-2.29-11n.tar.gz>.
10. M. Vranješ, S. Rimac-Drlje, and D. Žaga: Objective Video Quality Metrics. 49th International Symposium ELMAR-2007 focused on Mobile Multimedia, PROCEEDINGS ELMAR-2007, Zadar, Hrvatska (2007).

Author Index

A

Abhishek Mishra, [81](#)
Aditi Sharan, [567](#)
Aditya Sindhu, [273](#)
Ahuja, Mini Singh, [401](#)
Akki, C.B., [391](#)
Amita Sharma, [31](#)
Anbuudayasankar, S.P., [467](#)
Anjana Sharma, [177](#)
Anshul Gupta, [137](#)
Anubhuti Garg, [39](#)
Arjan Singh, [521](#)
Aruna Chakraborty, [487](#)
Ashutosh Gupta, [81](#)
Ashwani Kumar, [443](#)
Attlee Gamundani, [617](#)
Atul Bansal, [105](#)
Aulakh, Inderdeep Kaur, [319](#)
Avinash Gupta, [459](#)

B

Badjatya, Monika, [647](#)
Balram Swami, [265](#)
Bansal, R.K., [371](#)
Banyal, Rohitash Kumar, [155](#)
Bendre, M.R., [165](#)
Bhanodia, Praveen, [647](#)
Bharavi Mishra, [39](#)
Bhatt, Yogesh Chandra, [311](#)
Bhondekar, Amol P., [319](#)
Binod Kumar, [293](#)
Bishnoi, Lal Chand, [657](#)

C

Chakraborty, Aruna, [577](#)
Chandana, [477](#)
Chandraprabha Charan, [89](#), [147](#)
Charoo, Majid Ahmad, [637](#)

Chaudhary, Ajay, [499](#)
Chintan Bhatt, [343](#)
Choudhary, Surendra Singh, [499](#)

D

Deepika Sharma, [189](#)
Deepshikha Sharma, [155](#)
Deolia, Vinay Kumar, [105](#)
Dillen, Nicole Belinda, [577](#)
Dishant Mittal, [599](#)
Divya Saini, [559](#)

G

Garvit Bansal, [39](#)
Gaurav Bharadwaj, [47](#), [239](#)
Geetika Vyas, [31](#)
Geet Kalani, [229](#)
Ginika Mahajan, [411](#)
Gireesh Dixit, [301](#)
Gopal, [283](#)
Gowri, S., [611](#)
Gujar, Chandra Prakash, [647](#)
Gupta, Vikas, [531](#)

H

Hardik Molia, [351](#)
Hippolyte Muyingi, [617](#)
Hiren Kotadiya, [71](#)
Hudedagaddi, Deepthi P., [599](#)

I

Iti Sharma, [155](#), [559](#)

J

Jain, A.K., [301](#)
Jat, Dharm Singh, [657](#)
Jatinder Singh, [401](#)
Jaya Srivastava, [567](#)

Jibitesh Mishra, 359

Judy, M.V., 541

Juhi Sharma, 117

Jyoti Narwade, 293

Jyoti Pareek, 587

K

Kalai Lakshmi, T.R., 611

Kalyani, Vijay Laxmi, 89, 147

Karanbir Singh, 255

Kavita Kumari, 319

Khare, A., 301

Khehra, Baljit Singh, 521

Kiran Aseri, 47, 239

Kothari, D.P., 381

Kumawat, Naresh K., 229

Kuri, Manoj, 499

Kushwaha, Dharmender Singh, 459

L

Lokesh Agarwal, 301

Luv Sharma, 247

M

Maheshwary Shikha, 63

Mahnot Neha, 63

Maitri Jhaveri, 587

Malay Bhayani, 343

Mangesh Bedekar, 221

Manish Kumar, 443

Manish Mathuria, 477

Manoj Singh, 559

Meenu Dave, 127

Mehra Rekha, 63

Mehul Patel, 343

Mercy Bere-Chitauru, 617

Mishra, Durgesh Kumar, 637, 647

Mohammed Shehu, 627

Monika Gupta, 273

Monika Kunwal, 47, 239

Mugdha Gupta, 39

N

Nair, Prashant R., 467

Nakul Sharma, 55

Namrata Singh, 551

Neha Bansal, 105

Neha Solanki, 21

Nidhi Malik, 567

Nobert Jere, 627

P

Pandey, K.K., 301

Parmeet Kaur, 521

Pawanesh Abrol, 177, 189

Pharwaha, Amar Partap Singh, 521

Pinky Kumawat, 229

Pooja Pathak, 105

Popat, Kalpesh A., 351

Prabhat Thakur, 11

Prakash, S., 391

Prasanth Yalla, 55

Prashant Panse, 127

Pratipalsinh Zala, 71

Preet Kiran, 283

Priyabrat Dash, 359

Priyanka Sharma, 351, 433

Purvi Ramanuj, 199

R

Rachana Yadav, 95

Rachit Patel, 11

Raja, 411

Rajawat, Sharmishtha, 499

Rajesh Purohit, 21

Ravindar Singh, 265

Rishemjit Kaur, 319

Ritesh Kumar, 319

S

Sachin Wandkar, 311

Saksham Gulati, 333

Sakshi Kaushal, 255

Sandeep Joshi, 1

Sandeep Yadav, 95

Sandhu, K.S., 443

Saniya Zahoor, 221

Sanjay Agrawal, 137, 551

Sanjay Bhandari, 71

Sankalp Arora, 207

Sapna Katiyar, 11

Sarosh Umar, M., 423

Sasmita Pani, 359

Sathiyavathi, R., 611

Saurabh Bhardwaj, 283

Savina Bansal, 371

Saxena, Akash, 531

Sethi, Kamal Kumar, 647

Shadreck Chitauru, 617

Shah, J.S., 199

Shah, Kush, 513

Shaila Agrawal, 487

Shalki Sharma, 137

Shifali Kalra, 207

Shilpi Sharma, 333

Shivam Upadhyay, 89, 147

Shruti Mittal, 319

Shubhi Jain, 247

Smriti Srivastava, [283](#)
Sodhi, J.S., [333](#)
Soni, Bhanu Pratap, [531](#)
Sreeja Ashok, [541](#)
Sumeet Kaur, [371](#)
Sunita, [47](#), [95](#)
Supriya, B.N., [391](#)
Surendra Yadav, [477](#)
Swaleha Saeed, [423](#)

T

Tarun Shrimali, [127](#)
Thool, R.C., [165](#)
Thool, V.R., [165](#)

Tiwari, Pradeep Kumar, [1](#)
Tripathy, B.K., [381](#), [599](#)

V

Varad Vishwarupe, [221](#)
Vigneshwari, S., [611](#)
Vikas Bajpai, [39](#)
Vinod Mane, [221](#)
Vishwakarma, H.R., [381](#)

Y

Yash Gupta, [487](#)
Yogesh Rathi, [433](#)